

Comparing Apples to Oranges? Doing UX Work across Time and Space

Andrew Darby and Kineret Ben-Knaan

University of Miami, USA

Introduction

Usability testing is, famously, an iterative process. You test something, you make changes based upon the results, you test again. The recent website redesign process for the University of Miami Libraries began with an extensive “Discovery and Content Analysis” phase, which was the foundation for over a year’s worth of testing across the subsequent stages of the project. Each set of tests was designed to assess and/or improve the performance of specific features of the site and made use of a diverse set of methodologies and tools. While there was no static set of questions or tasks that appeared in all tests, those which performed well were removed from subsequent testing, while those which performed poorly continued on.

After almost 18 months of work, we had tested 75 questions or tasks, using seven different methodologies. We had written up 24 discovery reports and generated over 1,200 box files, which included meeting minutes, notes, raw data, pictures, Excel files, and more. We also had data in Google Forms, in an Optimal Workshop account, email discussions, printed mockups, and presentations to stakeholders. A data management strategy became necessary.

In this paper, we will outline our planning and discovery process, explore some sample activities to highlight the tools and techniques (and the heterogeneous data sources), and discuss how we tried to get a handle on all this data for future UX work.

Planning and Discovery

There is a temptation to do a website redesign just because you have not done one in a while; your head tells you that the purpose of the website is to provide useful information as seamlessly as possible, but your heart and your eyes start to wander. With these warring considerations in mind, we wrote up a redesign proposal for the University of Miami Libraries’ Administrative Leadership Team and laid out a number of answers to the question, “Why now?” Some were technical (a desire for a more agile and modern underlying architecture), some were practical (with a number of new applications coming online, we would need to be doing interface work anyway), and some were aesthetic (“modernize and prettify”).¹ A further overall reason was the fact that we were a substantially different organization from when the last site was launched: we had a new dean, a new Digital Strategies unit, a new Communications Team, an (under construction) Learning Commons, a new home for our special collections, and more in the works.

The proposal also laid out the team, timeline, and phases. The team was actually three teams, a newly established Website Redesign Advisory Team, the already-in-place User Experience (UX) Team, and the Web and Application Development department. The timeline was...ambitious. And the project was broken into four phases: Discovery and Content Analysis; Plan and Design; Build Out; and Post-Launch Assessment.²

Discovery Activities

The Discovery and Content Analysis phase focused on the current website and existing information sources as a foundation for the work moving forward. A series of activities were performed and then written up in reports with recommendations and notes on the process. These reports were then summarized in a master discovery report, which was shared with the Libraries’ Administrative Leadership Team, and meant to serve as a reference document for the work moving forward.³

Figure 1: Discovery Activities

Audits	Research Topics	Analytics	Other Data Sources	Focus Groups, Surveys, Tests
<ul style="list-style-type: none"> * Accessibility * Information architecture * Interface * Peer website review * WordPress audit * Page speed * Server analytics 	<ul style="list-style-type: none"> * HTTPS only * Advanced caching 	<ul style="list-style-type: none"> * Google Analytics * Heat mapping 	<ul style="list-style-type: none"> * Faculty interviews * University web guidelines * University Roadmap * Website comments 	<ul style="list-style-type: none"> * Staff survey * Departmental focus groups * Card sorting * One-on-one tests with students

These discovery activities can be grouped into five rough categories: Audits; Research Topics; Analytics; Other Data Sources; and Focus Groups, Surveys and Tests. The individual reports are available online, but it might be instructive to single out some reports for the purposes of this paper.⁴ **Audits**, as you might expect, were measuring the current site by some metric: the **Accessibility Audit**, for example, evaluated the site according to WCAG 2.0 AA web accessibility standards. This report found that the site met most web accessibility criteria, but also found a couple of points that needed remediation (mostly related to the search forms and keyboard accessibility).⁵ This sort of audit provides a clear assessment rubric, which can be rerun after the site is complete. The same might be said of the **Page Speed Analysis**, which used two different online tools to measure how quickly our pages rendered in both mobile and desktop environments. In this case, we compared our results to those of our self-identified peer institutions; these peer institutions provided a convenient random-but-not-random set of comparables.⁶ Some of the other audits were more informational and had less assessment potential. The **Interface Audit**, for example, revealed that we had a lot of mismatched buttons and forms and colours that had built up over time.

Research Topics were areas we knew we needed to investigate for the new site; for example, moving to an HTTPS-only site, which was a separate sub-project that ended up being completed before launching the new site. **Analytics** were data sources that did not require specific research or testing, just gathering results provided by a third party tool. Google Analytics had been running over the years, and so we reviewed the previous fiscal year’s results (and sometimes longer timeframes) and made a series of recommendations based on what we saw. Heat mapping, using the Crazy Egg software, was another tool we used for a bird’s eye view of how users interacted with the site.

Other Data Sources included information that others had compiled; we reviewed campus documents from the web communications team as well as the overall university’s “vision” document, the Roadmap for the Future; a series of faculty interviews that had been done by members of our Learning and Research Services unit; and then the online comments which users had submitted via a module of the SubjectsPlus software (<http://www.subjectsplus.com>).

Finally, we had more standard UX activities (**Focus Groups, Surveys, Tests**), which might be divided into those done with internal and external users. We began with internal users, and the purpose of these activities was twofold: to get some data on what our internal users wanted with the new site, but also more importantly to get them involved in the process. MailChimp (<https://mailchimp.com/>) was used to send out

an email directing users to the survey itself (Google Forms), and later used to send two reminder messages to the libraries' Listserv. Out of approximately 120 recipients, we had 58 responses. We also did focus groups with individual departments, which took a lot of time, but again was a good way of getting internal buy-in as well as generating ideas. We were also able to get perspectives from staff members outside of the public service departments.

The primary focus of the redesign of an external-facing website, however, should be with the end users. In the case of a large academic library, the primary (but not sole) user groups are students (undergraduate and graduate), faculty, staff, and alumni. The largest and easiest group to work with is students, and so, as part of the discovery phase, we did two types of testing with students: card sorting and task-based one-on-one user tests. The former test focused on the existing information architecture, specifically the menu system, while the latter tested a broad range of common website activities.

UX Methodologies

There are a large and growing number of methodologies or techniques commonly used in user experience research, which range from the simple (an online first click test or a one-question survey) to the complex (an ethnographic field study).⁷ Like many libraries, we do our UX work under resource constraints (both time and money), and this narrows the field of possible tests you might run: tests need to be cheap, and they need to minimize the amount of staff resources used to complete them. Having said this, even a relatively modest test takes up a fair amount of time, when you factor in planning, deployment, and analysis.

Our goal was to continue UX work at every stage of the overall process, so we used what we had learned in the Planning and Discovery phase to seed the testing of the next stages of the project. Working with this base set of questions or tasks, we selected a UX methodology that matched what we were working on in that phase of the project. For example, as we worked on the menu navigation in the early stages of Plan and Design, card sorts were a natural fit. We knew areas from our menus that were problematic, and we tried out alternate ideas. These results were used to generate questions for subsequent tests, which might use a different methodology. For instance, we did "first click" tests on home page prototypes, where we might include a question that tested the navigation, but other questions that tested the search box, the library hours, and other elements of the screen. These tests in turn would generate ideas for what we would test at the next stage of the design.

Getting a handle on everything we had tested and wanted to test becomes complex over time, so we adopted two different strategies to better communicate, manage, and assess our usability work for the overall project.

1. **Storytelling or Scenario Tracking**, where we retrospectively track different tasks over time and across tools. This is more of a communication method, internally and with our stakeholders.
2. **Data Management** approach, which involved archiving and collecting all data and insights to build our UX "database" with a visualization component. This gave us a high-level overview of the entire project and the ability to assess the website redesign project as a whole.

Storytelling or Scenario Tracking

As promised, we want to track how different user needs were tested and addressed across the testing and deployment period. We will step through two user needs (or tasks), taking a look at the tools along the way.

Library Hours

The importance of library hours came up regularly in our user research. Google Analytics is one of the easiest and most obvious places to start, and if you already use the service, the data is essentially in hand. During our discovery phase, we looked at results from the 2016 fiscal year, and saw that the hours page for the main library was the third most popular page on the current site, after the overall home page and the Music Library's home page. When we compared the 2016 results with those of the previous fiscal year, we noticed

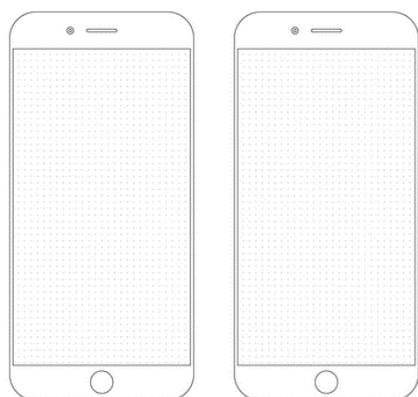
that, while the hits for many of the pages had gone down (uh oh), the raw numbers for the library hours page were virtually identical, and thus, in a sense, had gone up (as a percentage of hits).

So, nothing to do? Except, if you look at the source of the hits to the hours page, a whopping 74% of its hits are direct from Google. On the one hand, this is still good: people are finding our hours, even if they are not using our site to do so. On the other hand, this does not help us with the hours placement in our mockups. Beyond the analytics, we repeatedly heard and saw that the hours were important to our users.

Tool: Mini Design Sprints

During the design phase of our process, we ran “mini design sprints” with groups of users. A “design sprint” is a process that generally takes a week, where a group of designers and non-designers gets together to find a solution to a design problem.⁸ On day one, they outline the problem; they make sketches on day two; they review the sketches and vote on a possible solution on day three; the designers create a low-fidelity mockup on day four; and, on day five, the prototype is tested with users and feedback is collected. We modified this process to create “mini design sprints.” We had two variants. In one set of tests, we met with groups of internal users—for example, representatives of all the branch libraries together. We brought everyone into a room with a large table and a projector, and handed out pieces of paper with either an empty mobile or desktop template on it.

Figure 2. Sample Mobile Wireframes from SneakPeekIt.com



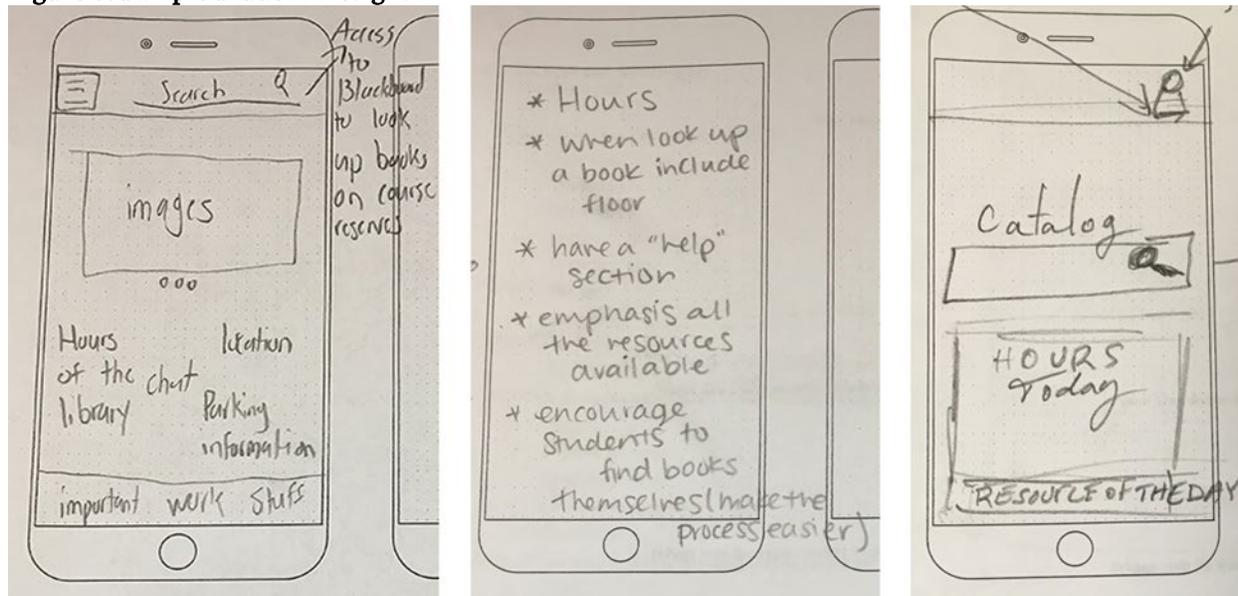
We asked users to complete two tasks, one which required a desktop sketch and another which required a mobile sketch. For example: “Design your branch home page in mobile—show the most important content/functions for you or your patrons in the visible area of the screen. You can test out more than one idea.” The users were given seven minutes to draw (after some experimentation, this seemed like a good amount of time), and then we took a picture of their drawing(s), uploaded them to Box (online file sharing software), and brought them up on the screen. Each user then walked us through their design. Next, we brought up on the screen a series of recommendations (with the UX test provenance) and asked if this recommendation would make them modify their design. For example: “Highlight subject guides and liaisons (subject specialists) on the main page of the website (*Faculty Interviews, Staff Survey*).” We found that framing the conversation in terms of our UX research was a good way to focus the discussion on areas we were interested in exploring (while also reminding stakeholders of the other work going on). Posting the sketches on Box also made it possible for people to teleconference in to the exercise.

These mini design sprints took about an hour and worked well, but the logistics of doing the same with faculty and students were daunting, so we tried something different. For faculty, we met with them individually in their offices and did not ask the follow-up questions based on previous UX work. For students, we set up a guerilla usability testing table just inside the entrance to the library and attracted participants with free coffee. The 17 undergraduate participants completed the sprint separately at one of

two stations at our table and were given the same prompt: “Design the Libraries” home page in mobile—show the most important areas for YOU in the visible area of the screen.”

Twelve of our 17 sketchers included library hours on the home page, making it one of the most consistent elements. Other consistent elements included **search** (15 of 17, and one of the two without a search box had a note “when look up a book include floor”), **room reservations** (8 of 17), and **log in/my account** (7/17). Sketches ranged from just a few bullet points jotted in the space to ones that created a layout.

Figure 3: Sample Student Designs



Hours are clearly popular with users, and thus figured prominently in student designs. By summer of 2017, we had three high-fidelity mockups in hand, and wanted to test which of these designs (or which elements from each design) were most successful for popular user tasks. One of these tasks was trying to find hours for a branch library.

Tool: Chalkmark

We chose to do “first click” tests using Chalkmark, an online tool from Optimal Workshop (<https://www.optimalworkshop.com/>) that allows you to test where on a prototype a user clicks. You upload a flat image, create your tasks, and then assign hotspots on the image that will count as correct clicks. It is possible to simply disseminate a URL for this sort of test, but we chose to set up a table in the main library again, with coffee as an inducement. In August 2017, we posed the same six tasks against the three different prototypes and cycled users through the three mockups. The facilitators did not do anything more than solicit users and pour coffee—the users completed the tasks on their own on a library-supplied laptop.

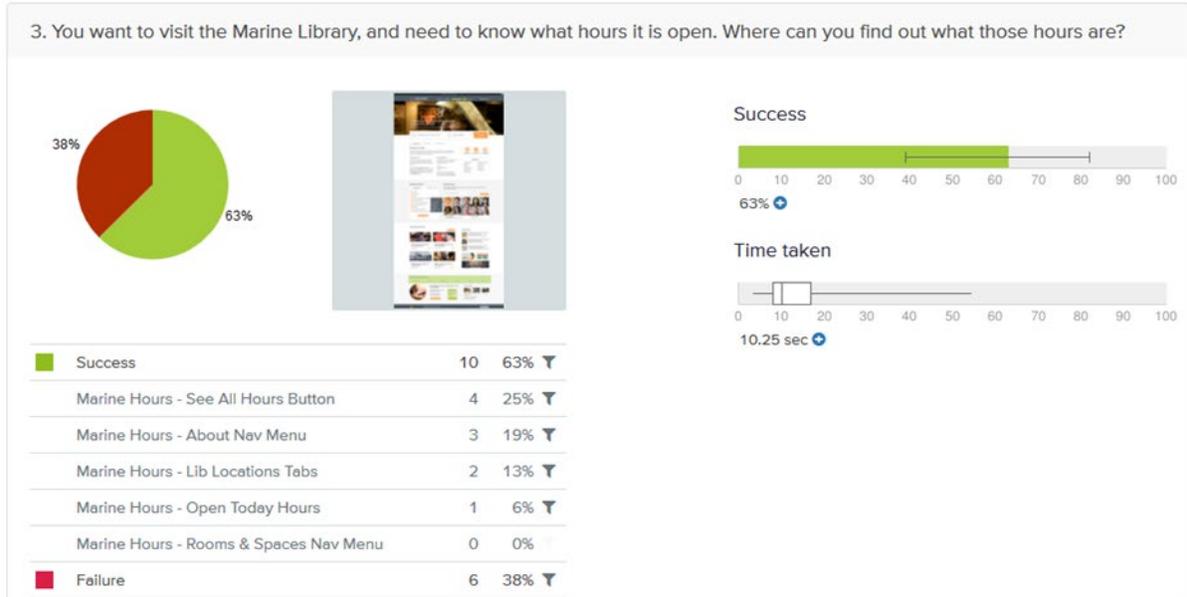
Specific to our story, we included one task about the Marine and Atmospheric Science Library’s hours: “You want to visit the Marine Library, and need to know what hours it is open. Where can you find out what those hours are?” Because the three designs were different, each had slightly different zones labelled as success.

The results for this particular question were not great. The designs had success rates of 47 percent, 50 percent, and 63 percent. The most successful design had a dedicated “Library Locations and Hours” band at the bottom of the page, an element that made its way into the final design. Since we were wondering if people would actually scroll down to see the content “below the fold,” we were pleased to see that users were finding this dedicated area. On the other hand, we were also testing our top-level navigation, and (unsurprisingly, with the power of hindsight) moving the links to our branches into a generic **Rooms &**

Spaces category was not a great success. Only 19 percent of our users found it in the navigation, and then it was under **About**. Wrong answer!

As for the tool, the results display in Chalkmark is pretty great. There is a simple gauge of success—the amount of time it took to complete the task, and then which of the designated “right answer” hot spots were clicked.

Figure 4: Task Results from Chalkmark



Later in the fall, we wanted to do a test of our revised information architecture. We had run some card sorting tests earlier in the year (which are off-topic for our current story) and wanted to test what we hoped was a close-to-final set of navigational categories and subcategories. This time, we turned to another tool from the Optimal Workshop suite: Treejack.

Tool: Treejack

In Treejack, you create your navigational structure in the software with the categories and nested subcategories, and then the user is presented with clickable links that chart possible paths through the navigational tree (see Figure 5). The software tracks the paths the user takes, including any times they switch back and try over.

Figure 5: Sample Task Presented to User in Treejack

Task 3 of 9

You want to visit the Marine & Atmospheric Science Library, and need to know what hours it is open.
Where can you find out what those hours are?



We ran the study for a week in October, with the following top-level navigational categories: **Services**, **Research**, **Rooms & Spaces**, **Libraries & Collections**, **About**, and **My Account**. Eighty-one people completed the study. For our Marine Library Hours question, we had a 98% success rate, with a 79% directness rate—that is, the percentage of users who went straight to the “correct” category of **Libraries & Collections**. One nice feature of Treejack is its “pietree” visualization.

Figure 6: Pietree Visualization (Detail)

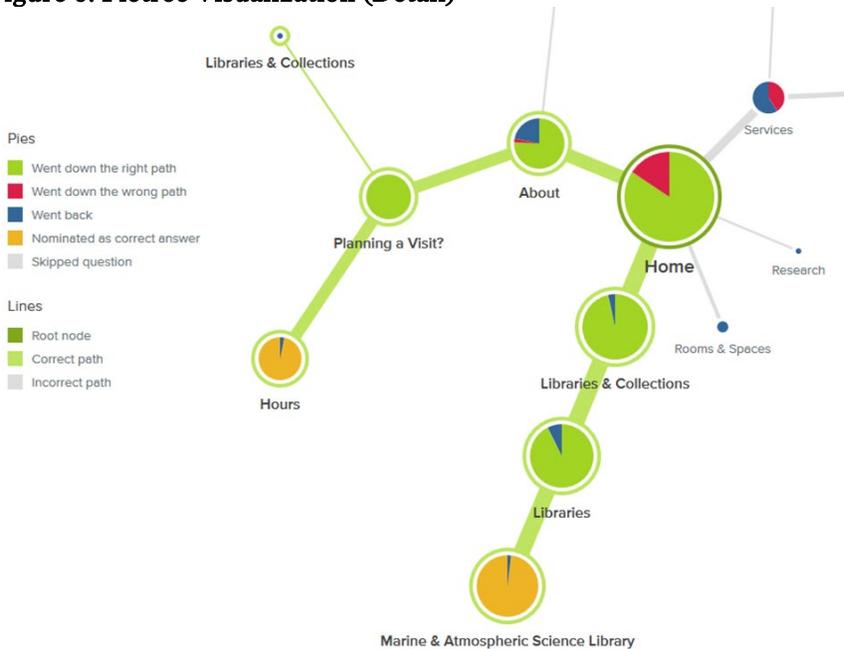
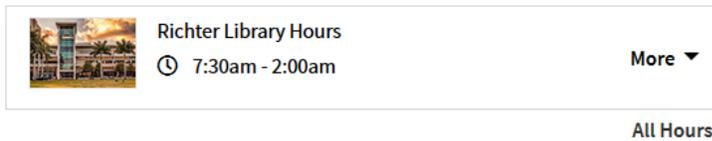


Figure 6 has the sort of visualization you would like to see: users mostly took the desired path (libraries and collections > libraries > marine), but the other path (about > planning a visit > hours) was equally desirable. At this point, we felt that we had a good handle on getting people to library hours on the main page and on a representative branch page. However, we also wanted to see how our library hours widget was working: this

was a little dropdown box that highlighted the particular library's hours, but also allowed one to access the other branch/collection hours (see figure 7).

Figure 7: Library Hours Widget



In October of 2018, we ran a one-on-one test in the main library, with a table set up just inside the front entrance doors and with a \$5 Starbucks gift card as the reward. One of the questions was a variant of our earlier questions: “You want to make sure the Marine Library is open before you drive over there. Can you find out the hours for the Marine Library?” We found that users took every possible route to the correct information: some went immediately to the widget and clicked “more” to view the Marine hours (2 of 7); others (2 of 7) scrolled all the way down and found it in the footer band; one user navigated to the Marine home page and looked on the widget there (which defaulted to the Marine hours); and then two found the hours listed directly in the **Libraries and Collections** mega menu. One of the users even said that she usually Googled the library hours, bringing us back to some of the findings from Google Analytics.

Project Data Management Approach

The importance of archiving past usability findings in a shareable way was noted in several articles published by Jakob Nielsen.⁹ Nielsen also stated that “the best usability reports are learning tools that help form a shared understanding across the team.” Following these best practices, our usability reports have been systematically published in a shareable Box folder. We have also collected notes, informal reports, and raw data. However, after more than a year, we had done a lot of tests, had a lot of data, and it was sometimes challenging to keep everything straight. While it was easy to compare results from one test to another, and trends were very clear when the same metrics were used, assessing our long-term project as a whole was challenging. As we have outlined, we used different tools and software and had different purposes and metrics. It was time-consuming to review a year and a half's worth of findings, and it was impossible to connect the dots or unravel hidden patterns by just reading the reports' recommendations. We wanted to maximize the value and insight that can be gained only through a high-level overview of the entire project.

Tool: Master Question List

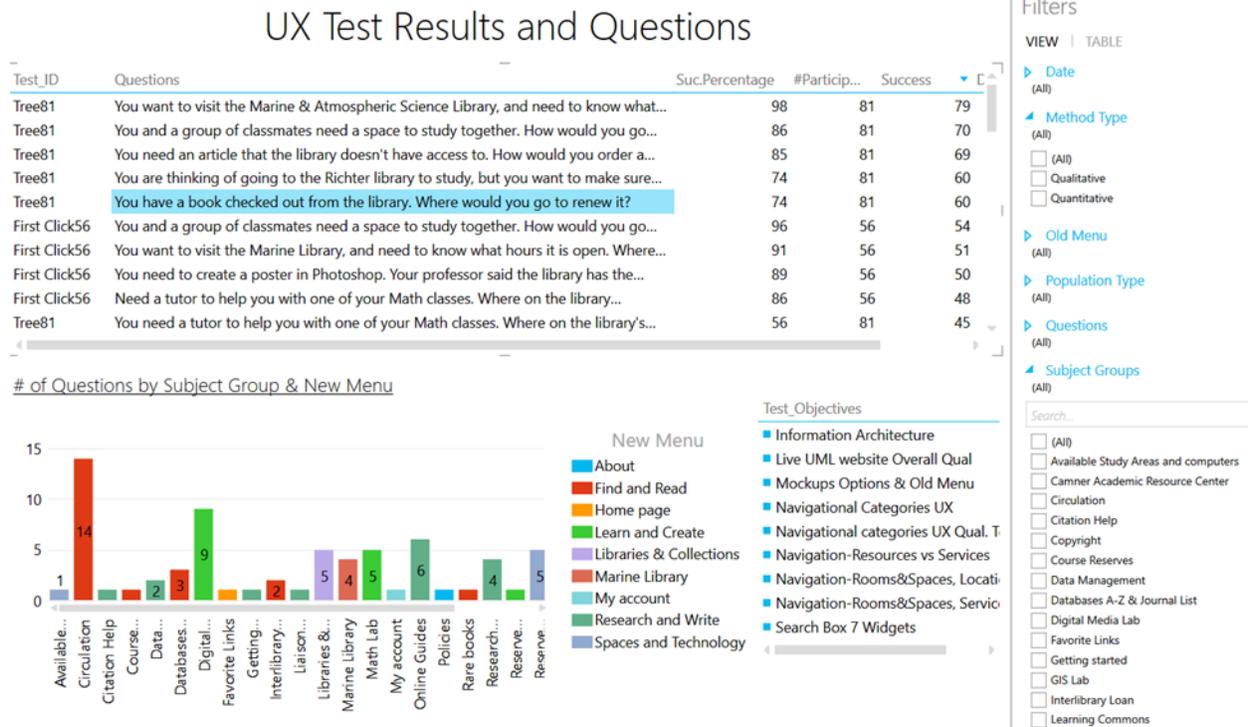
To this end, we used Excel to compile a master spreadsheet of our tests and activities, and then built out visualizations to explore the data. The master spreadsheet consists of the following tables and components:

1. **Test List:** Includes descriptive information about each test and its objectives.
2. **Menus and Subject Groups List:** Includes subject groups, which are types of services or resources (e.g., circulation, course reserve, policies, writing center) derived from the card sort tests. This table also includes the labels of the new and old top-level website navigation menus.
3. **Question List:** Includes all questions asked from all tests, both qualitative and quantitative; test ID and subject group names; and scores. If the results were quantitative, the scores were taken from the tool that was used for the test. Qualitative answers were scored by performance—meaning the UX team had decided in advance a scoring criteria for what was a correct answer based on the test objectives. The full qualitative test answers were stored in a separate Excel file.
4. **Card Sort:** Includes complete card sorting test data sessions and test ID.

Tool: Master Question List—Visualization Component

The visualization component was created with Excel’s Power View and Data Model.¹⁰ Power View helps to create interactive charts, graphs, and other visualizations inside an Excel workbook. The data model piece allows one to integrate data from multiple tables and to build a relational data source inside an Excel workbook. The reasoning behind choosing this tool over competitors like Microsoft Power BI was simple: it is free, easy to share, and confidential data can be protected. The Excel file was accessible to the UX team, and Power View updated automatically when additional data was added to the worksheet as we completed additional tests.

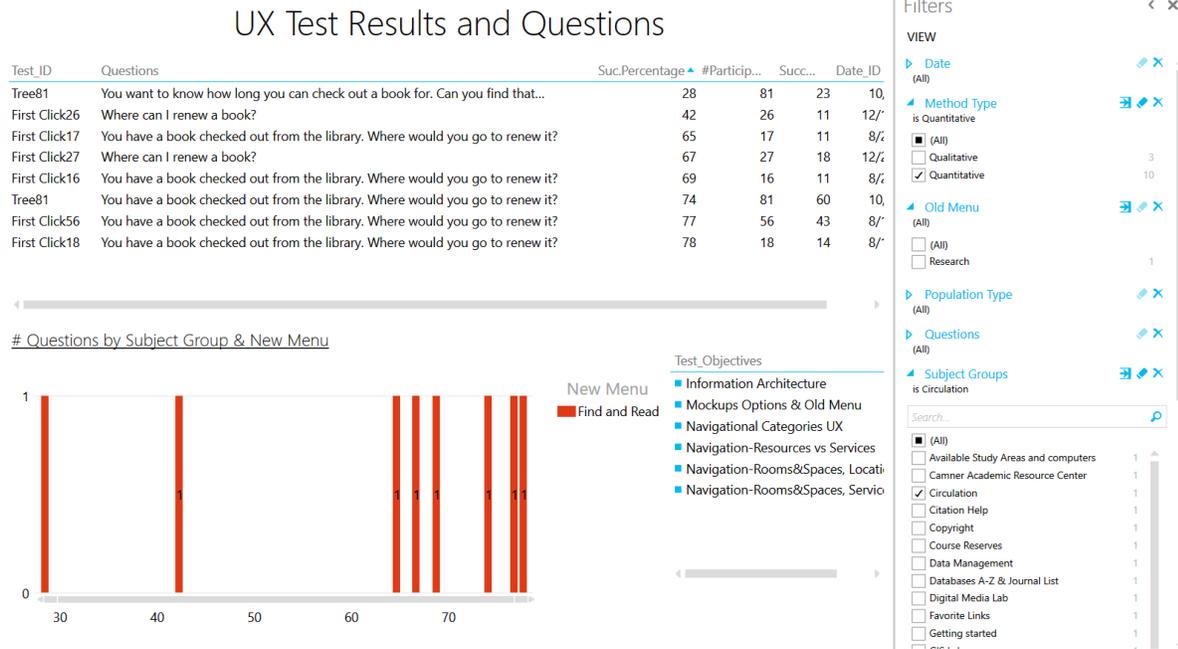
Figure 8: Master Question List and the Visualization Component



In the article “Archiving Usability Reports,” Nielsen singles out the power of performing a meta-analysis on cross-project data to gain insights that transcend individual projects.¹¹ He also recommends aggregating scores, success rates, task times, and satisfaction rates across multiple studies, first and foremost to answer the most basic question: “Are we getting better or worse?”

In addition to getting a bird’s eye view, we are able to use the Master Question List to drill down and answer more specific questions. For instance, we can use the filter option to limit to questions that had been coded as “Circulation,” and which took place at any point in the project (see figure 9). We can then easily see the different questions used and if there was variant phrasing; the testing methodology used; how many participants each test had; and crucially examine the relative success rates. We can further filter and compare the questions by population type (e.g., staff, students, or faculty) and the broad type of methodology (qualitative or quantitative).

Figure 9: Master Question List and the Visualization Component Filtered by Subject Group and Method Type



Creating the master question list spreadsheet and visualization enabled us to map the project activity and results. It should be noted that this was not a standalone decision-making tool, but more of a directional tool; it enabled us to see tests, scenarios, and questions in conjunction with one another. It helped us track success rate over time and prioritize what needs to be tested now. When we discovered interesting patterns that needed further investigation, we retrieved the relevant data and revisited our reports and documentation.

Conclusion

With hindsight, we would have done some things differently, but insofar as speed, iteration, and agility are fundamental aspects of UX work, it is hard to imagine a process where you knew everything you wanted to investigate from the outset. It was easier in the Discovery Phase to chart out a course, because you were working with a static set of information—the site as it existed at a particular time. As a result, we felt the Discovery Phase went well, but could have been less comprehensive. As is always the case with assessment, you need to think about what you are going to do with the data, not just collect data because you can.

Once we started testing out new ideas, especially prototypes, we did not have a whole course charted out—we saw problems, or now had new portions of the site in a testable state, and so we tested. We fixed what we could, or tried fixes, and tested again. The methodology changed, and the tests themselves were sometimes tests—for instance, early on we had a set of tasks that were relatively discrete, and later on started doing ones that were intended to flow as a whole imaginary research scenario. This meant the wording of tasks might make assumptions about a student (they are taking a class on Hamlet) that could impact the success rate. One thing we will definitely change moving forward is thinking more about how we score answers as right or wrong, whether it is binary or a point scale or something else, since that can really skew the results when they are looked at later.

Finally, once you have collected the data, make sure you actually use it! This is where setting up something like the master question list provides an invaluable bird's eye view of the process, so you can easily identify the pain points and where you need to do more work.

—Copyright 2019 Andrew Darby and Kineret Ben-Knaan

Endnotes

1. Andrew Darby, "Website Redesign Proposal," University of Miami Libraries, last modified October 10, 2016, <https://miami.box.com/s/1qxcg89kiuyt32k7w2rn29yx46jcmgul6>.
2. The final section consisted of a single, vague sentence: "Compare against benchmarks from Discovery phase."
3. Andrew Darby, "Discovery Overview Report," University of Miami Libraries, last modified September 28, 2018, <https://miami.app.box.com/file/138268067564>.
4. "Discovery Reports," University of Miami Libraries, accessed October 10, 2018, <https://miami.box.com/v/library-web-redesign>.
5. Charles B. Roberts, "Accessibility Audit Discovery Report," University of Miami Libraries, last modified November 27, 2016, <http://miami.box.com/s/tvz970mgqe8hfd1f3orsphoezxmuxyze>.
6. The ten self-identified peer institutions include: Brandeis University, Carnegie Mellon University, Case Western Reserve University, Emory University, New York University, University of Rochester, Syracuse University, Tulane University, University of Southern California, and Vanderbilt University. The most clear use of these peer institutions was for our Peer Website Review, which was a heuristic website evaluation. A heuristic evaluation "involves having a small set of evaluators examine the interface and judge its compliance with recognized usability principles (the "Heuristics" by Nielsen)." Unfortunately, we had not read this particular article at the time and our heuristic analysis ended up being over-engineered and an incredible time sink. It is the only case where we chose not to make our original report publicly available. Jakob Nielsen, "10 Heuristics for User Interface Design," Nielsen Norman Group, last modified January 1, 1995, <https://www.nngroup.com/articles/ten-usability-heuristics/>.
7. Christian Rohrer, "When to Use Which User-Experience Research Methods," Nielsen Norman Group, last modified October 12, 2014, <https://www.nngroup.com/articles/which-ux-research-methods/>; Raluca Budiu, "Quantitative vs. Qualitative Usability Testing," Nielsen Norman Group, last modified October 1, 2017, <https://www.nngroup.com/articles/quant-vs-qual/>.
8. Jake Knapp, John Zeratsky, and Braden Kowitz, *Sprint: How to Solve Big Problems and Test New Ideas in Just Five Days* (New York, NY: Simon & Schuster, 2016).
9. Jakob Nielsen, "Archiving Usability Reports," Nielsen Norman Group, last modified June 13, 2005, <https://www.nngroup.com/articles/archiving-usability-reports/>; Jakob Nielsen, "Formal Usability Reports vs. Quick Findings," Nielsen Norman Group, last modified April 25, 2005, <https://www.nngroup.com/articles/formal-vs-quick-usability-reports/>.
10. Microsoft, "Power View—Overview and Learning," Microsoft Office Help and Training, Office Support, accessed August 10, 2018, <https://support.office.com/en-us/article/power-view-overview-and-learning-5380e429-3ee0-4be2-97b7-64d7930020b6>; Microsoft, "Create a Data Model in Excel," Microsoft Office Help and Training, Office Support, accessed August 10, 2018, <https://support.office.com/en-us/article/create-a-data-model-in-excel-87e7a54c-87dc-488e-9410-5c75dbcb0f7b>.
11. Jakob Nielsen, "Archiving Usability Reports," Nielsen Norman Group, last modified June 13, 2005, <https://www.nngroup.com/articles/archiving-usability-reports/>.