

Lindsay Guarnieri, Tracey Kry, and Emily Porter-Fyke

# The eyes have it

## Using eye-tracking to evaluate a library website

**W**hen Western New England University announced its intentions to switch over its entire website from a legacy home-grown system to a brand new CMS, we were faced with moving all content on the library's website from one platform to another over the course of a summer. We needed to make our content fit into a strict new design scheme, but also wanted to take full advantage of the switch and use it as an opportunity to make our content work even better for our students.

To determine how successfully students were able to navigate the new library website, we partnered with our engineering department to conduct a usability study using eye-tracking software. In addition to useful information about how students use the website, we also learned a great deal about conducting research and working with outside partners. Through sharing our experience, we hope that anyone interested in conducting their own usability study will come away with tips, ideas, and pitfalls to avoid.

### Homepage of the D'Amour Library website

#### *Designing the study*

In order to keep the study manageable, we decided to determine exactly what we wanted to know about how students use the website, and what we didn't want to know. We wanted to answer these questions:

- Do students read the material presented or simply scan it?

- Are the buttons on the homepage confusing?
  - These buttons were designed with Flash to animate when moused over.
- Is there too much library jargon for students to translate?
- Is it clear where they need to go to accomplish their goals?

We did not want to evaluate:

- Our instruction program
  - At the time, each first-year student received research instruction for two class periods in the fall and two in the spring, which included being taught how to use the library website. We felt that we needed participants with a baseline of no library instruction in order to evaluate only the website.
- Website aesthetics
- Anything but the website
  - We didn't want to test students' use of databases, LibGuides, or other offshoots of the website over which we had little or no power of design.

### Design: Creating tasks

Our partner in the engineering department

---

Lindsay Guarnieri is former head of access services and electronic resources at Western New England University, email: lindsayguarnieri@gmail.com, Tracey Kry is archives and emerging technologies librarian at Western New England University, email: theresa.kry@wne.edu, and Emily Porter-Fyke, formerly of Western New England University, is now research and instruction librarian at Fairfield University, email: emilyporтерfyke@gmail.com

© 2020 Lindsay Guarnieri, Tracey Kry, and Emily Porter-Fyke

wanted to test both student use of the website and how the students learned to use the website as they used it. To that end, we needed to create specific, quick, and repeatable tasks for the students to complete as they tested the website, as well as a baseline of how easy the task should be (i.e., how many clicks it took us—power-users of the website—to complete). Our partner asked that at least two of the tasks be repeatable because that addressed his interest in learning if the students repeated a task, would their methods for completion differ as they learned more about the website? Additionally, our partner wanted some “impossible tasks,” tasks that simply could not be completed, such as finding course reserves for a professor who didn’t exist. This was another component of the “learning” aspect our partner was studying—if they came across an impossible task (it was not obvious to them whether it was achievable or not), how long would they try to finish that task, and what strategies would they employ before giving up?

With these requirements in mind, we created tasks based on what we considered to be the most common things a student might need to find on the website, such as:

- a book,
- other resources (article, etc.),
- reference contact info,
- a certain librarian’s contact info,
- course reserves, or
- library blog.

We created 17 tasks of various iterations of these common goals, most of them re-

peatable and two of them impossible. For the most part, we didn’t have any difficulty thinking of tasks for participants to accomplish. It was, however, difficult to keep to



Homepage of the D’Amour Library website.

our resolution that we did not want to test anything but the website. In order to evaluate whether the participants successfully accomplished a task, the task had to have a clear end. This made our desire not to test offshoots of the website difficult because so much of our functionality depended on other platforms, such as our discovery service (“Find IT! @ D’Amour Library”), which is powered by EBSCO, or our database landing page and Research Guides, which come from LibGuides. We compromised by designating the “end” of the task as simply locating the area in which the task goal would be, such as scrolling to the book

record or other resource.

### Execution: Recruiting students

As mentioned above, we decided that to get a real idea of how user-friendly the website was, we would need to perform usability testing with students who had had no previous experience with the site. We took a two-pronged approach, with the library recruiting through flyers, social media, and tabling, and our partner recruiting in his department mostly through teaching and word-of-mouth. The biggest draw was the raffle for students who signed up to participate in testing to win one of four \$25 gift cards to a local panini restaurant.

Most of the final participants came to us from our partner’s recruitment efforts. This meant that most of our participants were

engineering students, interested in the chance to use the eye-tracker firsthand. While this was great for our final participant count, we also wonder if it might have led to skewed results.

In the end, we recruited nine participants, all first-years who had had no or very little prior library website instruction. We had to make some small concession to this requirement as the semester wore on, and we worked with the students' busy schedules to arrange appointments for them.

### **Execution: Testing sessions**

The testing sessions were completed under the supervision of our engineering partner, as they were held in the engineering department using their equipment. For those who are unfamiliar with eye-tracking technology, the device consists of two wearable parts: the glasses and the recorder. There is also accompanying software. The glasses used by our engineering department (the Tobii Pro Glasses 2) have two cameras per eye, which record the movements of the wearer's pupils to see where they're looking, and a scene camera, which records what the wearer is viewing.

They are able to track the movements of the wearer's eyes by illuminating the eye, which creates reflections, and the glint of the light in the cornea and pupil is used to calculate where the person is directing their gaze. The glasses, recorder, and software match that up to the surroundings recorded with the scene camera, so you don't have to be an expert to use them.

So in the case of our usability testing, the person doing the test put the glasses on, calibrated them by gazing at a fixed point for about 15 seconds, and then proceeded through the test wearing the glasses. The glasses recorded both what the subjects were looking at and how their pupils moved to gaze at it, resulting in data such as heat maps and gaze plots.

Once all sessions were completed, our partner shared video and audio recordings of the sessions, complete with eye movements, as well as data pulled from the eye-tracking

software. In addition to heat maps and gaze plots, the software also provided the amount of time taken, in seconds, for each participant to complete each individual task.

### **Execution: Quantitative vs. qualitative data**

Next, we analyzed both the data and recordings. We quickly learned that numbers could only tell us so much. Some tasks took longer because of design issues with the website, while other tasks took longer simply because they involved more steps. To fix this, we used our baseline time for each task and calculated how much over the baseline each participant took to complete the task. We did the same for click counts.

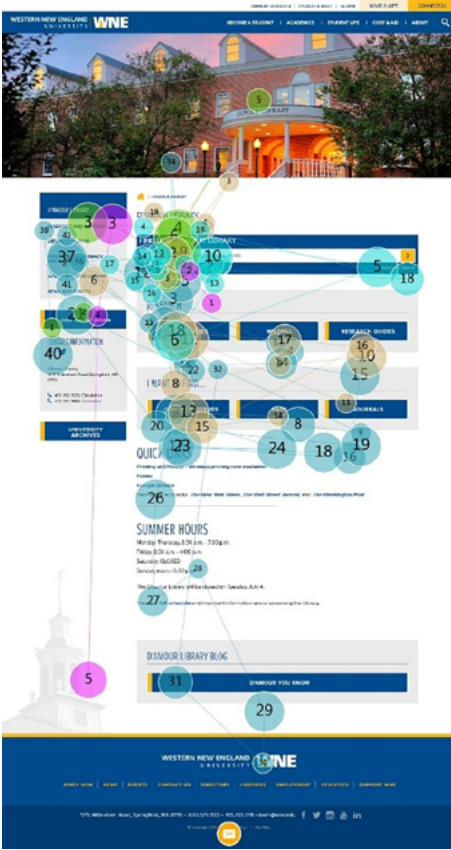
In addition to looking at quantitative data, we needed to investigate the more nuanced actions of the students that couldn't be defined by a number. We spent a good deal of time with the videos, examining each thoroughly to answer questions the numbers couldn't tell us, such as:

- Did they complete the task successfully?
- Did they take the optimal path?
- Did they scroll more or less than expected?
- What did they do when confused?
- Did they read content or just scan?<sup>1</sup>

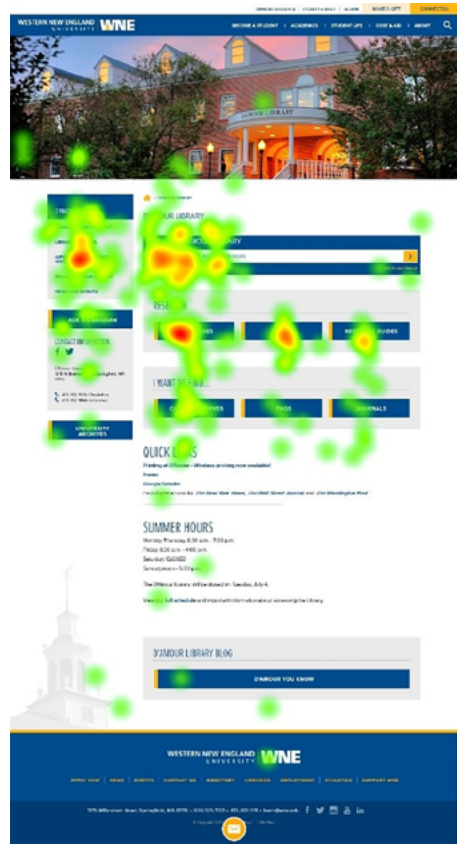
### **Results**

Our results were not surprising, but they did confirm many suspicions and concerns we had not only about the functionality of our website, but also about student habits and tendencies. In regards to the website, the biggest actionable takeaways from our study were that the lefthand navigation of the page was looked to and used more often than any other portion of the page, items below the fold (the top half of the page) often go unnoticed, and if there is a search box, it will be used, sometimes regardless of the search box's intent.

Regarding individual tasks, the task that proved to be most difficult for students to



Aggregate gaze plot of all test participants.



Aggregate heat map of all test participants.

Task	Optimal Time	Average Time	Overage time (average - optimal)	Optimal click count	Average Click Count	Extra clicks (average - optimal)
Find the course reserve list for Professor Beagle	16.69	96.554	79.864	2	8.1	6.1
Find the course reserve list for Professor Clark	13.516	29.426	15.91	2	3.78	1.78
Find the course reserve list for Professor Doe*	n/a**	n/a**	n/a**	n/a**	7.3	n/a**
<b>All course reserve tasks</b>	<b>15.103</b>	<b>62.99</b>	<b>47.887</b>			<b>3.94</b>
Find the Library's "Reserve a study room" form*	n/a**	n/a**	n/a**	n/a**	8.89	n/a**
Find the Library's "Make a Suggestion" form	13.125	30.461	17.336	2	2.13	0.13
<b>All forms tasks</b>	<b>13.125</b>	<b>30.461</b>	<b>17.366</b>			<b>0.13</b>
Find the Library Director's email address	15.094	82.9	67.806	2	4.3	2.3
Find the Staff Assistant's email address	12.75	56.212	43.462	2	3.6	1.6
<b>All contact info tasks</b>	<b>13.922</b>	<b>69.556</b>	<b>55.634</b>			<b>1.95</b>
Find the call number for the book "The Grapes of Wrath"	17.493	73.806	56.313	2	4.89	2.89
Find the call number for the book "To Kill a Mockingbird"	48.766	81.092	32.326	2	8.3	6.3
Find the call number for the book "The Hobbit"	18.367	97.089	78.722	2	7.55	7.56
<b>All call number tasks</b>	<b>20.209</b>	<b>83</b>	<b>62.791</b>			<b>5.58</b>
Find resources about "global warming"	17.024	83.777	66.753	2	8	6
Find resources about "The Civil War"	17.406	50.408	33.002	2	5.63	3.63
Find resources about "artificial intelligence"	19.54	51.631	32.091	2	5.5	3.5
<b>All resource tasks</b>	<b>17.87</b>	<b>61.939</b>	<b>44.069</b>			<b>4.38</b>
Find the Library's hours	9.493	22.13	12.637	0	0.56	0.56
Find the Library blog	10	59.18	49.18	0	3.22	3.22
Find information about printing at D'Amour	10.633	36.69	26.057	1	2.22	1.22
Find a way to contact the Reference Librarians to ask a question or get help	9.977	33.745	26.713	1	1.33	0.33
<b>All homepage tasks</b>	<b>10.026</b>	<b>37.936</b>	<b>27.91</b>			<b>1.33</b>

\*impossible task

\*\*impossible tasks' times and clicks were not included so as not to skew results

Chart of task times and clicks.

complete was finding a course reserve list for a specific professor. And because of the repeating tasks, we could identify the most difficult group of tasks, which was finding call number information. By counting clicks we determined that students took the optimal path to information more often than not when looking for information on the homepage (hours, contacts, the blog, etc.), and took the least optimal path when looking for call numbers.

## Lessons learned

In future usability testing, impossible tasks will be omitted. In most cases, they resulted only in frustration, and in some cases the randomization of tasks placed them at the top of the list. Having an impossible task for the first task seemed to hurt the participants' confidence and impacted the rest of the session.

While repeated tasks can be informative when determining the “learnability” of a website, we suspect that they skewed the results of our usability study. In many cases, we found that once students learned a method to find something on the website, they used it every time a similar task appeared, even if it was not an efficient method (i.e., they were clicking much more than they needed to). Students were occasionally confused, thought that it was the same question, or they assumed it was a mistake and did not try as hard as they might have otherwise to complete the new, repeated task. If repeated tasks are to be included in a study, we would recommend making this clear to subjects before testing begins.

We also suspect that there may have been issues with jargon being (unintentionally) built into tasks. For example, because we were interested in whether students would notice and use the FIND IT! search box, some of the tasks asked the participant to find “resources” on a topic, rather than specifying that they find a book or find an article. Multiple participants were confused by the term resources, and were unsure whether they had successfully completed these tasks. Another result we questioned was whether any of the students would have been able to complete the task to

find course reserves if we had not specifically used the jargon course reserves, which was how the button was labeled on the homepage. We encourage anyone planning a usability study to closely interrogate the use of jargon when designing tasks.

Jargon in the website itself also proved to be a problem. Our library, like so many others, named the OPAC with an acronym—WILDPAC. Perhaps when OPACs were still the go-to resource for research, this practice worked, as the vaguely named resource was the only choice. Now, with so many places to search, it is overlooked because it is an unknown—the name tells users nothing about what it does. Not a single participant used the OPAC during their session, but the heat maps and gaze plots show that students did in fact look at it. This tells us that it was unused not because it was inaccessible or difficult to find on the page, but because participants didn't know what it was.

Additionally, nearly all of the students at some point gave up on using the organization of the site as the means to finding the goal of a task, and instead began methodically clicking through each link in the left-hand navigation and scanning for relevant wording before moving on. It was unclear whether this was because there were too many tasks or because students were frustrated and confused by repeated and impossible tasks.

In analyzing and applying what we learned throughout the study, it is difficult to compartmentalize issues related purely to the website, and those related to instruction, and the ways in which we teach students how to use the website. But the very nature of libraries and the work we do on a daily basis creates this unavoidable overlap. We hope to take what we learned and improve our website where we can, and continue to look at the bigger picture of how to best serve our students in whatever way they need us.

## Note

1. Supplemental resources including the template scoring sheet are available at <http://bit.ly/eyes-have-it-resources>. 🦋