The Unobtrusive “Usability Test”: Creating Measurable Goals to Evaluate a Website

Tabatha Farney

Academic libraries strive to present user-centric web interfaces that will easily guide users to the information and resources they need. However, website effectiveness is typically based on usability testing that takes place during a website redesign process and not as a standard, continuous practice. This can be problematic because an academic library’s intended audience is constantly changing as students come and go every semester. Libraries need to constantly test the effectiveness of their website to ensure it is still meeting user needs. Traditional usability tests, such as user surveys or focus groups, will help but are not always feasible to implement on a regular basis due to time constraints and limited resources.

A simple, innovative solution to this problem is to monitor website usage statistics using web analytics tools, such as Google Analytics, Piwik, or AWStats. These unobtrusive instruments collect and analyze website usage data that reveal how users are actually interacting with a library’s website without disturbing the users. While these tools cannot replace traditional usability tests that share the users’ thought processes as they interact with a site, they do provide constant, easily obtained use data that a library can use at any time to improve its website’s functionality. This paper demonstrates how libraries can apply the concepts of usability testing to the reviewing and reporting of website usage data collected by a web analytics tool. It highlights how the web services librarian for the Kraemer Family Library at the University of Colorado Colorado Springs implemented this process and provides practical advice for libraries interested in adopting this practice.

Literature Review

There are many different definitions for usability, however not all are specific to website usability. Nielsen and Loranger define website usability as the ease of using a website, specifically in regards “to how quickly people can learn to use something, how efficient they are while using it, how memorable it is, how error-prone it is, and how much users like using it.” Rubin and Chisnell promote usability testing as the “process that employs people as testing participants who are representative of the target audience to evaluate the degree which a product meets specific usability criteria.” They describe a process that involves identifying desired website users and defining specific usability metrics to measure a website’s efficiency, effectiveness, user satisfaction, and user ability to learn how to use the site.

Usability testing often implies having participants perform tasks to evaluate the website based on usability metrics. However, there are several types of usability tests and each varies in design, purpose, and ideal time to implement them during the website development cycle. Summative testing, sometimes referred to as validation testing, takes place after the website is finished and validates that the site meets its desired goals based on set benchmarks. It often uses a large number of users and focuses on simple usability metrics such as success/failure rates on tasks or average time on task. Summative testing is the ideal candidate to use a web analytics tool because it offers a large data set based on actual interactions on a website and does not rely on self-reported data.

Tabatha Farney is the Web Services Librarian at the University of Colorado Colorado Springs, e-mail: tfarney@uccs.edu.
data reports traffic to a website as visits or visitors; it collects and reports visitor related data which can show how website users are interacting with a site. Common website metrics include visits, visitors, page views, duration of visit, page depth, and bounce rate. While there are many other metrics, these data points alone can reveal how much use a web page receives and how long visitors stay on a website. All of this data is collected unobtrusively using web analytics tools, such as the popular Google Analytics, that also helps create easy to read reports based on these metrics.

Libraries are actively using website usage data to aid the decision making process for redesigning websites, to identify and measure strategic website goals, and to validate the success of a redesign project. This paper is a continuation of the research that demonstrates how collecting usage data with web analytics tools can continuously monitor the usability of a library website.

Background
The Kraemer Family Library at the University of Colorado Colorado Springs is an academic library that supports nearly 10,000 undergraduate and graduate students. The library’s website is edited by several librarians comfortable using the university’s content management system, but the development and management of the website is handled solely by the web services librarian. The website received its latest major redesign in 2010 and has not been comprehensively evaluated since then. To prepare for migration to a new content management system, the web services librarian needed to evaluate the site’s current usability to determine if the site was still meetings its intended usability goals. With limited time and resources to test the site before the migration was scheduled, the web services librarian decided to use website usage data to evaluate the site’s usability. During this process, she found essential information that proved invaluable and was able to configure the library’s web analytics tool, Google Analytics, to regularly report this information so she could be updated on the site’s performance during and after the migration. As a result, she is able to use these same reports to continuously monitor the performance of the website based on usability goals and to fix potential usability errors as they arise.

Methodology
Before a library can use a web analytics tool in usability testing, the tool must be properly set up on the website to receive the most accurate usage data. Each web analytics tool has its own implementation process and may require customizations to track all the necessary data. Since this case study utilizes Google Analytics, this paper will discuss the necessary customizations the web services librarian implemented for this particular tool.

If Google Analytics is not already recording data for the website, the first step is to create an account and set up a profile for each website to be tracked. After a profile is created, it generates a unique tracking script that must be embedded into the HTML code of each web page on the site for the tool to accurately track data from the entire site. Google Analytics can only report data after the tracking script is in place. Other customizations to the Google Analytics profile or to the website are also required to optimize the tool’s data tracking ability. A popular, and often necessary, customization is tracking outbound link that direct visitors to another website.

Once Google Analytics is implemented, the next phase is to craft a usability test plan. The web services librarian loosely adapted Rubin and Chisnell’s usability test plan that includes identifying the usability test’s purpose and goals, research questions, targeted website users, and specific tasks.

Identifying the Usability Test’s Purpose and Goals
The intent of the usability test should be clearly defined because it outlines all other elements of the test including the goals and tasks. The purpose should be specific in nature. A purpose that simply states “improving the user experience” is far too broad and implies too many variables that overwhelms the test and will bury potentially useful outcomes. The purpose should directly relate to either the website’s strategic plan or to the original intent behind the website redesign.

For this case study, the purpose of the usability test was “to measure how effectively and efficiently the library’s website assists users in finding research materials online.” The goals of the study include:

- Measure the ease of accessing research databases that are located on the library’s website.
- Evaluate how users connect to additional research assistance.
- Improve users ability to find online research guides.
These goals directly connect to the website’s primary and secondary purposes identified in the library website’s strategic plan. Although additional goals could be identified, the web services librarian chose these goals because they fit the study’s intent and perceived need of library users.

**Writing the Research Questions**
Research questions are extensions of the test’s goals and should focus on concerns that impact the user experience. The web services librarian identified three main research questions:

- Can library website users access a research database quickly and efficiently using the library’s website?
- What benchmarks can be created to evaluate how users find research assistance on the library’s website?
- How do users find an online research guide on the library’s website?  

These research questions will drive the tasks selected for the test, so they should outline specific elements in the website that will be tested.

**Targeting the Users**
In traditional usability testing, selecting the participants for the study can be an arduous task because the participants need to be a representative sample of the website’s users. When using website use data, the default unit of metric is visitor. In web analytics terms, a visitor is not an individual person, but rather the web browser or device that the person uses to access the website. Hence, a person that accesses the library’s website from their laptop and then accesses the website via their mobile device is counted as two separate visitors on the website because the visit occurs on two different web browsers. Regardless of this multiplicity, visitors are an ideal target population because they are actual website users.

Google Analytics defaults to providing aggregated visitor data, but contain features that can segment usage data into more defined user groups. This includes, but is not limited to, comparing usage between new visitors and returning visitors which could be used to measure a website’s “learnability.” It is assumed that returning visitors will complete tasks more quickly or efficiently because they are already familiar with the site and learned how to use it from an earlier visit.

The web services librarian targeted three users groups: all visitors, visitors that completed the desired tasks, and new and returning visitors. Analyzing all visitors puts in perspective how often tasks were accomplished on the site, while focusing on the visitors that completed the desired tasks would provide insight on how those tasks were accomplished. New and returning visitor data was used to evaluate whether or not the website helped users learn how to use it—this was not part of the test’s research questions so is not a major focus of this case study.

**Selecting the Tasks**
A task is simply a desired action for the user to perform on a website. For in-person observational usability testing, a test moderator typically assigns the participant a task and then monitors and records how that participant attempts to complete the task. A sample task related to a library’s website could be to locate the library hours for a specific day or find a journal article on a suggested topic.

Tasks should also contain a benchmark that defines the successful completion of the criteria. This is especially true for tasks designed to be measured with website usage data because a completed task does not always imply it was a positive user experience. For example, if two participants were assigned the same task, but one took one minute to finish it while the other needed ten minutes to finish it, this could imply that the second participant found it more difficult (and probably frustrating) to complete the task. Benchmarks do not have to be precise, but they should be based on the user’s expectations.

Most task-based usability tests are limited to the number of desired tasks so that participants can complete the test in about an hour. There are no such restrictions when substituting web analytics for in-person usability testing, but it is best to prioritize the tasks based on the test’s research questions to keep this process manageable. Data analysis requires time, so more tasks equals greater time commitment for library staff.

Since this study recommends website usage data for usability testing, it is essential that the tasks identified can be measured by the use data. In other words, there has to be a metric for the variable being tested. A sample of viable tasks used in this case study include:

- Task 1: Find a research database within four web pages of the visit.
The Unobtrusive “Usability Test”

- Task 2: Navigate to an online research guide within four web pages of the visit.
- Task 3: In less than four minutes, find how to get more research assistance on the site.
- Task 4: Use a chat widget after being on a web page for longer than one minute.25

Each task has its own assigned benchmark. The first two tasks have the same benchmark of completing the task within four page views because this particular library’s website was designed so that users could access a majority of the site’s online resources within three clicks regardless of where they entered the website. Hence, those first two tasks are directly testing to see if the website performs as designed. The other two tasks use time as a benchmark. These benchmarks were designed based on Nielsen and Lortanger’s findings that users spend an average of 27 seconds on each web page26 and a total of 3 minutes and 49 seconds to complete a task.27 If users are on a web page longer than a minute or on the website longer than four minutes, it could imply that the user is engaged with the content or that the user is confused and requires more assistance.

Defining the Necessary Metrics and Reports

Google Analytics does not automatically generate usability specific reports or data because it is not designed to be a usability testing tool. However, the data collected in Google Analytics can be used to assess website usability by replacing traditional usability metrics with usage data for each assigned task. Typical usability metrics for summative type tests are total successes/failures for each task, the time it takes to complete a task, completion/success rates, and common used pathways to complete the task.28 All of these traditional usability metrics can be translated into either a metric or report found in Google Analytics.

The most essential web analytics metric for this study is the conversion, sometimes simply referred to as a “goal” in Google Analytics. In web analytics terms, a conversion is simply a desired action on a website.29 Whenever that desired action is performed on the website, a conversion is triggered and that conversion-related data is reported in a goal report. Google Analytics provides several goal report types, but this study focused on the goal overview report that displays the total number of conversions completed and conversion rate (the number of conversions divided by total visits) in addition to the reverse goal path report that outlines the last three steps visitors took before making the conversion. Goal reports are a custom feature that must be configured within the website’s profile before the goal report reflects any data.30

For this usability test, conversions are a direct match with task completion; therefore each task requires the configuration of a conversion which will generate a separate goal report for every conversion. However, conversions alone do not reveal if the task was successfully completed as specified by its benchmark. To find that data, each goal report needs a custom segment based on the benchmark criteria to filter the successful completions from the non-successful completions. For example, the first task’s conversion is clicking on a research database link (tracked as an event in this case study31) and the necessary custom segment determines the successful conversions that took place in visits with four page views or less from unsuccessful conversions by visits with more than four page views.

The web services librarian opted to display the time it takes to complete the task metric as two data points: a mean and a range. The mean provides the average time it takes to complete the task while the range put the data into perspective by displaying the various times it takes visitors to complete the task. The web services librarian used two different Google Analytics reports, audience overview report and the visitor engagement report, to find these data points. The audience overview report contains basic visitor data for the entire website, including the average time visitors are on the site. By applying a custom segment to show only visits where the particular conversion/task was completed, the web services librarian was able to find the average time it takes visitors to complete the task. The visitor engagement report displays the time on site data as a range by grouping visits by the length of time they spend on the site. While this method did not show the exact time on site for each visit, it does produce a useful range. The web services librarian applied the same custom segment to find the average time on site to this report to create the range of times it takes visitors to complete that task.

The completion/success rate was also separated into two data points. The web services librarian used the conversion rate for the completion rate because it demonstrates how often the task is completed on the website. The conversion rate is available in the goal overview report. The success rate is designed to prove
whether or not the task can be accomplished successfully on a website on a regular basis. This metric is determined by dividing the total successful conversions by the total number of conversions. Most usability tests set a benchmark to measure whether or not the task is successfully completed. The common success rate benchmark is between 66% to 70%, but the higher the success rate the more user-friendly the task. Success rates below 66% identify tasks that are not particularly user-friendly, and those tasks should be further analyzed to determine the common usage for visits with failed tasks to fix the usability errors.

The final usability metric is the commonly used pathway data that describes the paths visitors used to navigate through a site before completing a task. This information can be gathered in the reverse goal path report that displays the last three steps visitors made before completing a task. While this report is limited to the last three web pages visitors viewed, it still provides an easy method for determining the route a majority of visitors used to finish the task.

**Reporting the Findings**

The final phase in the usability test plan is to document how the test findings are reported. A major benefit of using a web analytics tool such as Google Analytics is that it can collect and analyze website use data to automatically generate a report. It is possible to create a goal report based on each task, create and apply the custom segments, and use other data.
reports as necessary, all within the Google Analytics tool. There is even an option to schedule an automatic email that exports the reports on a desired set schedule. This serves as a useful, scheduled notification that also provides all the necessary data for the preliminary usability analysis. Currently, the web services librarian set the scheduled email to monthly to analyze and report the data at the end of every month to identify errors related to a recent website migration to a new content management system. By June 2013, the reporting schedule will change to the end of each semester.

Although Google Analytics makes reporting data fairly easily, the web services librarian opted to take all the individual data points and combine them into one usability findings report. This allows her to analyze and report the data findings for successful and failed tasks in one document. For the failed tasks, it is essential to identify the source of the errors, rate the severity of the errors, and provide actionable recommendations to fix these errors. The web services librarian creates this information based on the data analysis and includes the findings in the regular usability report.

Results
Using Google Analytics to provide website use data for regular usability testing was found to be a seamless solution to keep track of the website's performance based on its usability criteria. After the initial customizations (setting up goal reports, creating custom segments, and ensuring all content was being properly tracked), the web services librarian was able to focus her attentions on analyzing and reporting the data. By using the scheduled email feature, she did not have to remember to go into the tool to pull the data—it was automatically sent to her email signaling that it was time to draft the report.

Example of the Data Analysis Process
Analyzing the data, identifying usability errors, and making recommendations to correct the errors are the most time consuming parts of the process. Table 1 contains the usability metrics for the first and second tasks gathered over a three month period. The first task has a success rate of nearly 67%—according to Nielsen this is an average success rate so the web services librarian opted to keep monitoring it without further error analysis. The second task’s success rate is extremely low at almost 40% and triggered the need for further error analysis by the web services librarian to identify the potential problems and solutions to the errors.

<table>
<thead>
<tr>
<th>Excerpt of Data from Usability Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Total Conversions</td>
</tr>
<tr>
<td>Total Successes</td>
</tr>
<tr>
<td>Total Failures</td>
</tr>
<tr>
<td>Success Rate</td>
</tr>
<tr>
<td>Average Time on Site</td>
</tr>
<tr>
<td>Visit Duration</td>
</tr>
<tr>
<td>0-10 seconds</td>
</tr>
<tr>
<td>11-30 seconds</td>
</tr>
<tr>
<td>31-60 seconds</td>
</tr>
<tr>
<td>61-180 seconds</td>
</tr>
<tr>
<td>181-600 seconds</td>
</tr>
<tr>
<td>601-1800 seconds</td>
</tr>
<tr>
<td>1801+ seconds</td>
</tr>
</tbody>
</table>

The first step in error analysis is to look at other related usability and website use metrics that could help identify the problem. For the second task, the web services librarian used the average time and range of times visitors took to complete the task and the visit duration range to see that website users were noticeably on the site much longer when searching for a research guide. On average visitors were on the site for over 15 minutes before finding a research guide—this seemed like an extensive amount of time since the average time on site for a regular visitor was under 9 minutes. Additionally, looking at the range data revealed that while 30% of the visitors took less than 30 seconds to find a guide, 41% needed longer than 10 minutes to find a guide.

Once the web services librarian understood the length of time it took for users to find a guide, she next analyzed the trends in how visitors found the guide using the commonly used pathways data shown in table 2. An estimated 60% of the visitors took a similar path where they entered the site at the library’s homepage, went to the guides listing web page, and then selected a specific research guide. This pathway
suggests these visitors knew exactly where to navigate to find a research guide which confirms the web services librarian’s anecdotal assumption that most research guides were class-specific guides that librarians direct users to in library instruction sessions. Yet, this pathway is under four page views, but it did not correlate with the task’s success completions or success rate. This is the result of tracking the research guides as “pageless views” in Google Analytics, which is no longer the preferred method for tracking outbound links. The web services fixed this tracking error for the next round of usability testing.

Aside the tracking issue, nearly 40% of visitors was still not completing the task within four page views according to the commonly used path data. The second most popular path had visitors navigating through various pages and eventually returning to the library’s homepage to find the guides page and then a research guide. The third path showed users were viewing the library’s homepage twice before successfully navigating to a research guide page. Both of these pathways suggest problems: visitors are not sure where to click on the library’s homepage to find the guides, and visitors are not aware they can use a dropdown navigation feature found on every page to easily connect to the research guides pages. While both of these are errors, their severity was noted as a “cosmetic problem” since they are not happening as frequently as originally assumed, and users can clearly overcome these hurdles to find a research guide. Overall, the web services librarian recommended fixing these errors by identifying the trends in the random library pages visitors were on before navigating to the library’s homepage to find the guides page—the intent was to focus on the top ten pages found in the first step of the pathway and make a “quick links” widget linking to the research guides page so visitors could find the necessary web pages in fewer page views with hopefully less frustration. The web services librarian will monitor the progress of the task in further usability testing to determine if these revisions assisted in improving the overall success rate.

Drawbacks to Using Website Use Data for a Usability Test

While this method proved extremely useful to the Kraemer Family Library, it was not without its imperfections. Yet, most of the issues which include the inability to see participants interacting with the website, to talk directly with participants, and to confirm participant motivation for their actions will also be found in any asynchronous remote usability testing. Website usage data demonstrates user interactions, not user motivations. It is difficult to tell why that visitor came to the library’s website—so it is not possible to accurately measure incomplete or failed tasks because a user who comes to the website for a research database, but never ends up finding it, is recorded as just another visit. There is no method to flag those visits without some type of online survey that asks website users to share their reason for visiting the library’s website as they enter or leave the site. If a library implements such a survey, it could potentially collect that data as a custom variable in Google Analytics which could make it possible to connect user motivation data with basic website usage data. Further research is required to test this possibility.

The other major caveat is that website usage data is information about interactions on a site and cannot provide data from a site’s non-users. The closest web analytics metric to simulate a non-user are bounces and bounce rates which indicate that visitors entered a website only to leave it right away. Farney and McHale describe this interaction similar to “a patron coming into the physical library and then turning

<table>
<thead>
<tr>
<th>Pathway</th>
<th>Completions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enter site -&gt; Library Homepage -&gt; Guides Page -&gt; Research Guide</td>
<td>1899</td>
</tr>
<tr>
<td>Random Library Page -&gt; Library Homepage -&gt; Guides Page -&gt; Research Guide</td>
<td>612</td>
</tr>
<tr>
<td>Library Homepage -&gt; Library Homepage -&gt; Guides Page -&gt; Research Guide</td>
<td>434</td>
</tr>
<tr>
<td>Library Homepage -&gt; Random Library Page -&gt; Guides Page -&gt; Research Guide</td>
<td>78</td>
</tr>
<tr>
<td>Not set -&gt; Enter site -&gt; Guides Page -&gt; Research Guide</td>
<td>39</td>
</tr>
<tr>
<td>Unique Paths (outliers)</td>
<td>16</td>
</tr>
<tr>
<td>Total Completions</td>
<td>3140</td>
</tr>
</tbody>
</table>

Aside the tracking issue, nearly 40% of visitors was still not completing the task within four page views according to the commonly used path data. The second most popular path had visitors navigating through various pages and eventually returning to the library’s homepage to find the guides page and then a research guide. The third path showed users were viewing the library’s homepage twice before successfully navigating to a research guide page. Both of these pathways suggest problems: visitors are not sure where to click on the library’s homepage to find the guides, and visitors are not aware they can use a dropdown navigation feature found on every page to easily connect to the research guides pages. While both of these are errors, their severity was noted as a “cosmetic problem” since they are not happening as frequently as originally assumed, and users can clearly overcome these hurdles to find a research guide. Overall, the web services librarian recommended fixing these errors by identifying the trends in the random library pages visitors were on before navigating to the library’s homepage to find the guides page—the intent was to focus on the top ten pages found in the first step of the pathway and make a “quick links” widget linking to the research guides page so visitors could find the necessary web pages in fewer page views with hopefully less frustration. The web services librarian will monitor the progress of the task in further usability testing to determine if these revisions assisted in improving the overall success rate.

Drawbacks to Using Website Use Data for a Usability Test

While this method proved extremely useful to the Kraemer Family Library, it was not without its imperfections. Yet, most of the issues which include the inability to see participants interacting with the website, to talk directly with participants, and to confirm participant motivation for their actions will also be found in any asynchronous remote usability testing. Website usage data demonstrates user interactions, not user motivations. It is difficult to tell why that visitor came to the library’s website—so it is not possible to accurately measure incomplete or failed tasks because a user who comes to the website for a research database, but never ends up finding it, is recorded as just another visit. There is no method to flag those visits without some type of online survey that asks website users to share their reason for visiting the library’s website as they enter or leave the site. If a library implements such a survey, it could potentially collect that data as a custom variable in Google Analytics which could make it possible to connect user motivation data with basic website usage data. Further research is required to test this possibility.

The other major caveat is that website usage data is information about interactions on a site and cannot provide data from a site’s non-users. The closest web analytics metric to simulate a non-user are bounces and bounce rates which indicate that visitors entered a website only to leave it right away. Farney and McHale describe this interaction similar to “a patron coming into the physical library and then turning
right around and walking out.” While not a positive action, it still requires the visitor to access the library’s website in order to be included in the data.

**Implications for Academic Libraries**

Despite its imperfections, applying the structure of a usability test to analyze website use data is a simple and effective strategy for regularly testing a library’s website with minimal time, commitment, and setup. Additionally, it does not disturb your website users because the data is captured instantly with each visit. Libraries that already implement other types of usability testing can apply elements of their usability test plans to this method by repurposing test goals/tasks so they can be measured using website usage data.

Aside from conversions and goal reports, there are other useful web analytics metrics and reports that can provide data to measure the user experience on a website. Analyzing the time to load site/page metric, use data related to 404 error pages, site search data, and any web page with a high bounce rate are just a few examples of data points that libraries can use to identify potential errors on a website that could cause user frustration. By taking the time to analyze website use data on a regular schedule, libraries can leverage the wealth of information it provides to implement revisions and improve the overall user experience for their sites.

**Conclusion**

Website use data and web analytics tools are just one component of a library's overall website usability plan, but they are instrumental in keeping web services librarians and administrators informed about how users approach their websites even after the formal usability testing period ends. The data is constantly being collected, so it can be regularly analyzed to identify new errors as new students enter the library each semester. It is a time saving solution that has the potential to bring a high return on investment if a library is committed to developing their web analytics usability plan and to monitoring and reporting the data. Configuring a web analytics tool to help with the data collection and automate the reporting process will allow for more time to focus on identifying potential usability problems and proposing actionable solutions to remedy the problems.

**Notes**

4. Ibid., 35-37.
6. Ibid., 20.
16. Ibid.
17. Excerpt from Kraemer Family Library's usability test planning document.
18. Ibid.
23. Ibid., 81.
25. Excerpt from Kraemer Family Library’s usability test planning document.
27. Ibid., 27.
34. Ibid., 261-263; and Barnum, *Usability Testing Essentials*, 263-264.
39. Ibid., 45.