

Performance Metrics

Zarla Ludin

Bentley College

4/7/2008

Introduction

The following analysis is of data gathered on three informational websites. We looked at the data in the website of interest, Site 1, in comparison to two other competitor sites: Site2, Site 3. The data collected on all three sites were from 300 participants performing a series of eight tasks to test the usability of the site. We designed the session based on the client's goals, which included testing the overall usability of Site 1, and comparing the success rates of tasks of Site 1 to Sites 2 and 3.

We randomly assigned participants one of the three websites (Site 1, 2 or 3) on which they would conduct the eight tasks. The participant demographics include age, household income, and number of visits to the site in which they were assigned. Three performance metrics were recorded during the test, which are task success (recorded in a binary system), time-on-task (in seconds), and the number of page views to complete the task.

From this analysis, we hope to offer some insight into the issues that may affect effectiveness, efficiency, and user satisfaction when using the site.

Methods and Analysis

A Microsoft Xcel spreadsheet automatically recorded the data as participants took part in the usability test. For the purposes of this evaluation, we will focus our attention on the three core performance metrics: task success, time-on-task, and page views on the site. In addition, we will examine an interesting correlation between task success and the number of visits participants made to the site.

Task success, as was mentioned above, is determined by scores 0 and 1. A score of 0 means failure and 1 means success. Because we tested three sites, we examined the percentage of success rates for each site. To achieve the average success rate, we simply added up all the instances of successful task completion (all the scores of 1) and compared them to one another. From this analysis, we can see an overview of which site had the highest percentage of successfully completed tasks.

Time-on-task is a way for us to evaluate just how much time a participant spent on a particular task. For the purposes of this analysis, we wanted to find the most problematic task for Site 1 participants. During the planning stages of this test, we determined that all tasks should take roughly the same amount of time, therefore analyzing the average time it took to complete a particular task will help us pinpoint which one (or more than one) task proved to be the most difficult on Site 1.

We will then look at the average number of page views for Page 8 per site. This will tell us how far participants came within each site, which may show us some insight into the participant's tolerance to stay on a particular site. Tolerance to remain on a site is an indicator of how useable and enjoyable the site is.

Finally, we will examine the correlation between the amount of site visits and task completion. For this analysis, we will examine in particular the time it takes to complete Task 8 and the corresponding amount of site visits each participant made.

Executive Summary

Site 1 is the most usable site from the three tested. Task success rate for Site 1 was 72.5%, which is higher than the overall average of 63.3%. Task 6 is the most problematic task—it took the longest (roughly 140 seconds to complete when the average time to complete a task was roughly 97 seconds) and had the lowest success rate for 2 of the 3 sites. Site 1 however had the best success rate with Task 6. We found that no significance exists with the number of Page 8 views; nearly all sites had the same amount of traffic

on Page 8. Finally, we found a negative correlation between the number of site visits and time-on-task for Task 8. That is, the more visits a participant made to the site, the less amount of time it took for them to complete Task 8.

Data reliability

As was mentioned above, we collected the data using an online study approach. The main advantages of this approach are that multiple participants can access the session at the same time, and that more participants can take part than in a traditional usability session in the lab. However, the yielded results from an online study are not as loaded with subtle information that can be observed through one-on-one human contact. Participants in online tests may not be as engaged in the study as well. In examining the raw data for this test, for example, it is clear that some participants carried out the test simply for the incentive, completing tasks in mere seconds. In addition, there are clear outliers in the raw data that have high task completion times. To remedy this, we have omitted the outliers for analysis. In addition, we used confidence levels in our analysis, which are the “I” shaped graphics in the charts.

Findings

Success Rates

To begin, we wanted to get the big picture of this enormous data set. Because we measured the success rate using a binary system, we were able to determine with which site’s participants were most successful. Using standard descriptive statistics, we calculated the percentage of successful participants per task per site. As was mentioned in the previous section, we have included confidence levels. For success rates, we used the Adjusted Wald confidence interval. Site 1 had around a 73% success rate, which is much higher than the overall average rate of 63%.

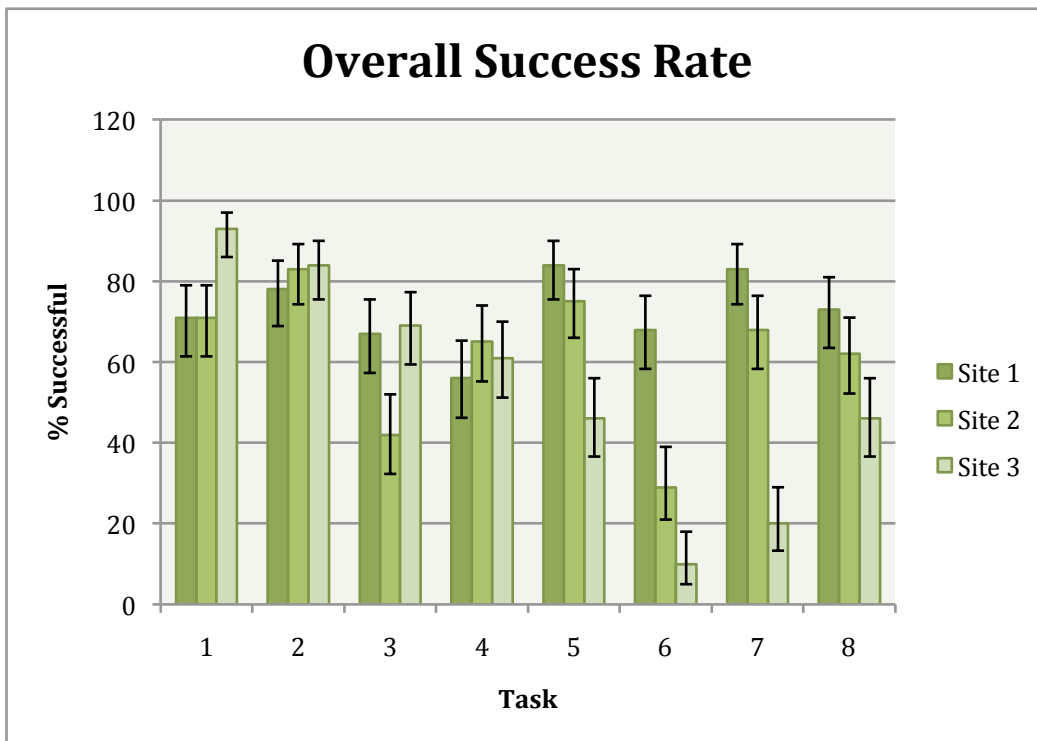


Figure 1. We determined the success level of each sites’ set of 8 tasks. Clearly, Site 1 had the most successful amount of tasks completed during the test.

We wanted to determine the success rate by task as well. To do this, we again used standard descriptive statistics when calculating each task's average success rate. From this, we found Task 6 to be the most problematic task, in that it was the least successful task. Only 36% of participants were able to complete Task 6, compared to the overall mean of 63%.

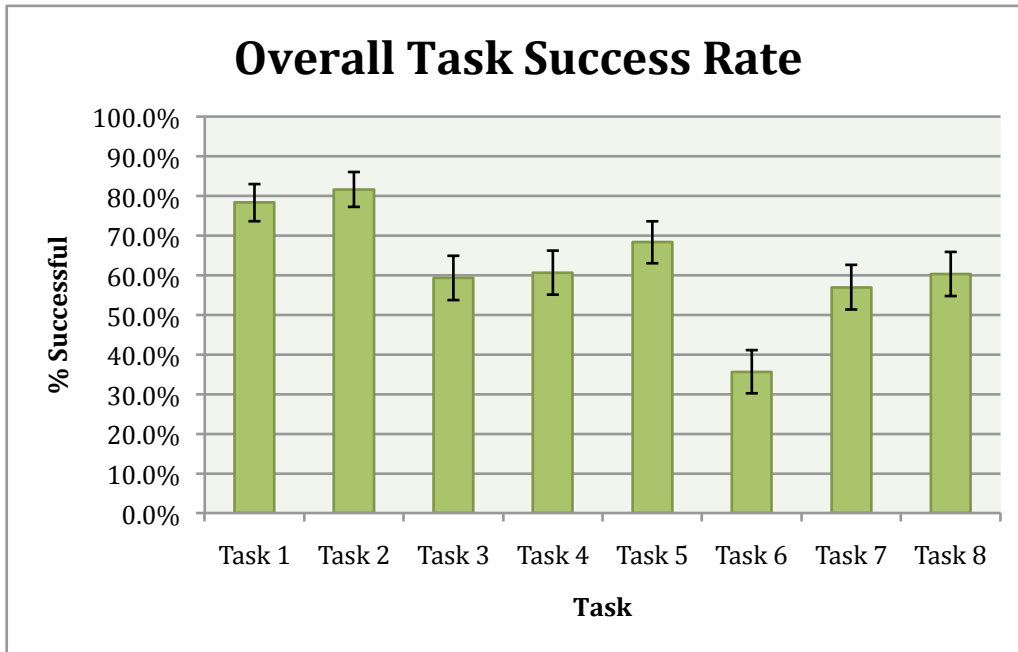


Figure 2. Task 6 was the least successful task. It was also the task that took the longest to complete.

Time-on-Task

Task 3 took the least amount of time to complete overall. This may be due to the fact that we had organized the tasks from easiest to most difficult, and participants had had some time to learn the site. Task 6 took the most time to complete among the sites. Because the long task time transcended all three sites, it may be that the task was just too difficult to complete. However, we suggest looking at the design of the corresponding pages to this task considering all tasks were predicted to take roughly the same amount of time to complete.

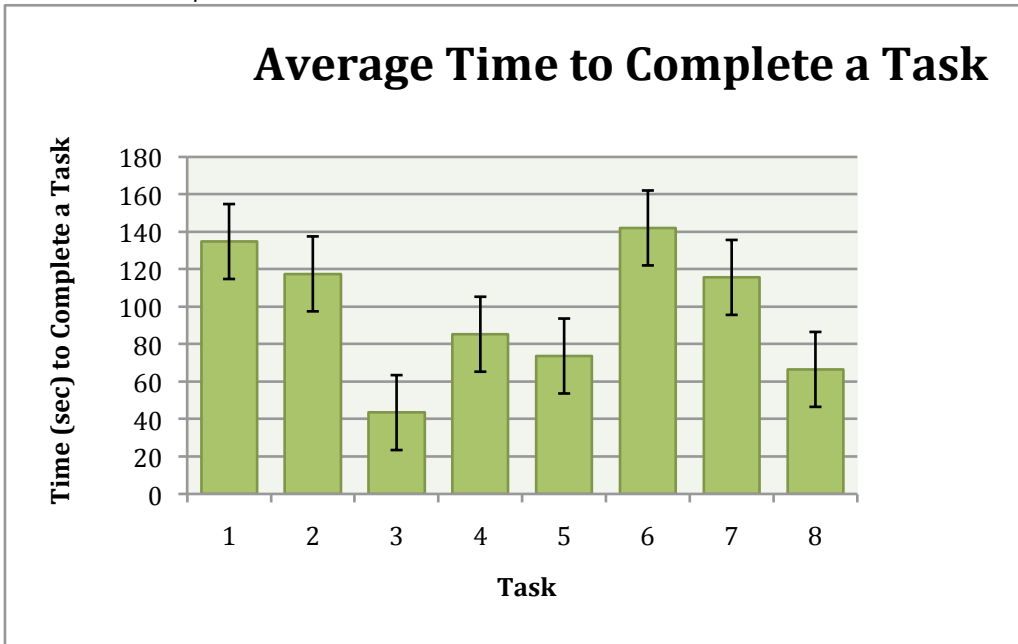


Figure 3. Task 3 took roughly 45 seconds to complete, while Task 6 took around 140 seconds. Considering all tasks were predicted to take around the same amount of time, this is a huge difference.

Page Views

We wanted to compare the number of views Page 8 received for each site. Using ANOVA, we wanted to see if any significant difference exists between the number of page views on each site. The F-value yielded was 1.45, which is below the F-Critical Value of 3.03 that needed to be achieved significant. We From this analysis, we can infer that all of the sites were usable enough to maintain user participation until the end of the session.

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	78.26	2	39.13	1.4467	0.237000646	3.02615
Within Groups	8033.2	297	27.04788			
Total	8111.5	299				

Figure 4. The F-Value did not exceed the F-Crit, therefore the number of page views between each site was not proven to be significantly different.

Task 8 Success and Number of Site Visits

To determine if the number of site visits affected the success of a task, we compared Task 8 times with the number of site visits. We simply used Xcel to visualize any relationship between the amounts of time it took a participant to complete Task 8 and looked for a relationship between those times and the participant’s visit number. We created a scatter plot to visualize this comparison. When we inserted a trend line, we found a negative correlation. The negative correlation means that the more times the participant visited the site, the amount of time it took them to complete the final task, Task 8, decreased. This may mean that participants are becoming well-acquainted with the site and have an easier time using it by the end of the session.

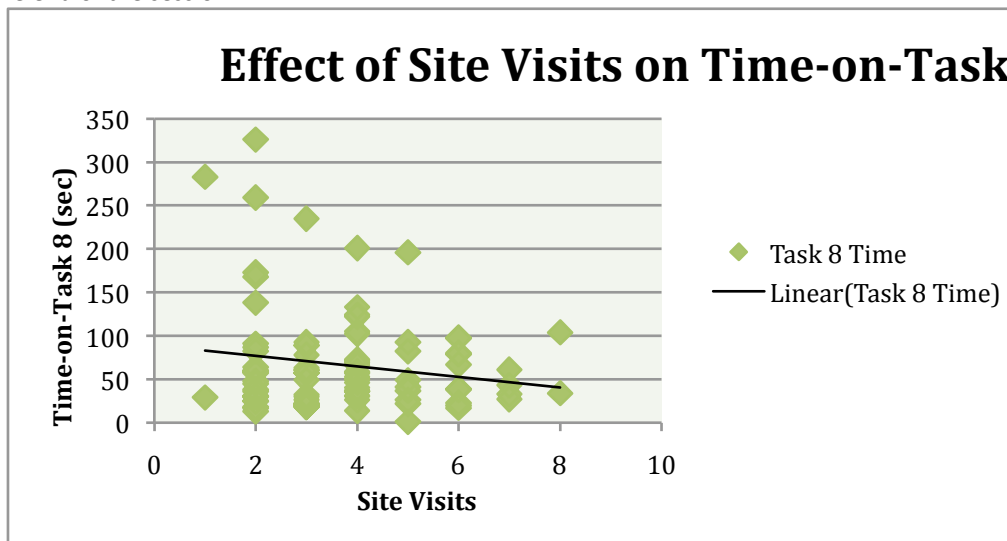


Figure 5. As the number of site visits increase, the time-on-task decreases.

Conclusions

Overall, the design and usability of Site 1 is more successful in comparison to its competitors, Sites 2 and 3. In addition, Site 1 participants were the most successful at completing tasks. The high success rates of Task 3 show that that page was very well-designed, and can be a source of good design when reevaluating the site. We also found that participants were engaged in the session until the very end, which can correspond to user's positive impression of the site. Finally, we determined that participants were able to learn how to use the site, which is good for efficiency and effectiveness of the site. We determined that participants were able to learn how to use the site by correlating the number of site visits to the Task 8 time-on-task time. We also omitted any analysis of individual participant cases from our report because we felt it would not contribute to the findings.

Although Site 1 is more usable than Sites 2 and 3, it can still be further improved to create a more satisfactory experience for the user. We suggest reexamining the pages involved in completing Task 6 for example. This proved to be the most problematic task across all three sites. However, this data expresses that Site 1 is doing well in comparison with its competitors.