

IBM “MANY EYES” USABILITY STUDY

Team Six:

Rod Myers

Dane Petersen

Jay Steele

Joe Wilkerson

December 2, 2008

EXECUTIVE SUMMARY

We asked three graduate students in Informatics at Indiana University to complete a series of scenarios and tasks on IBM's Many Eyes website. All were experienced computer and Internet users but novice users of the Many Eyes website, and our primary objective was to test the ease of learning to use the site. The users had little difficulty with most of the tasks and found it especially easy to publish visualizations and make Wordles. They particularly enjoyed making Wordles and exploring the related features. All users were frustrated by the registration and login process. They didn't understand that the IBM ID needed to be an e-mail address, and the error messages did not help them to resolve the problem. They also had difficulty finding existing visualizations of datasets, as well as some difficulty finding a targeted dataset.

Based on our analysis of the usability data, we conclude that the site is initially somewhat difficult to learn; however, once the users have made it through the registration process and explored the features of the site, they find it fairly easy to use. We recommend improving the registration and login process so that field requirements are clear and error messages are more specific and helpful. We also recommend modifying some page layouts so that navigation elements are visible and grouped to make choices clearer.

TABLE OF CONTENTS

Executive summary	1
Table of Contents	2
Introduction	3
Description of the test.....	3
Test objectives.....	3
Summary of user profile (target population)	3
Summary of test preparation	3
Testing Schedule	3
Session length	4
Test Setup.....	4
Description and rationales for the testing scenarios	4
Description and rationales of the tasks.....	5
Evaluation methods and results.....	5
Usability goals	5
Summary of evaluation strategies.....	5
Description of data analysis methods and selection rationales	5
Summary of data analysis	6
Classification of the severity of the problems and rationales.....	8
Screenshots from the testing to document findings.....	9
Usability problems and recommendations	11
Appendix	12
References.....	13

INTRODUCTION

DESCRIPTION OF THE TEST

We evaluated IBM's Many Eyes website (<http://manyeyes.alphaworks.ibm.com/manyeyes/>). The people behind Many Eyes are part of IBM's Collaborative User Experience (CUE) research group which studies computer-supported cooperative work. The purpose of Many Eyes is to facilitate social data analysis and visualization by allowing people to upload datasets, create visualizations of those data, add metadata, and share their work with others. A primary goal is to spur "discussion and collective insight about data."¹

TEST OBJECTIVES

Our primary objective was to test the ease of learning to use the Many Eyes website. In particular, we studied the ability of novice users to explore datasets and existing visualizations and to upload new datasets and create original visualizations.

SUMMARY OF USER PROFILE (TARGET POPULATION)

We chose university graduate students as our target population, our rationale being that they would be likely to use a website like Many Eyes to explore data gathered during their own research. We recruited three participants (two males and one female) from among graduate students in the Informatics program at Indiana University. All were between 23 and 35 years old. All three participants rated themselves as expert computer users and expert or near-expert Internet users. Two participants rated themselves as near-expert with data visualization tools (the meaning of which was prompted by the text "e.g. graphs and charts in Excel, Google Charts, etc."), while the third was near-novice. None of the participants had ever used the Many Eyes website.

SUMMARY OF TEST PREPARATION

TESTING SCHEDULE

We conducted all three of our usability tests on the afternoon of November 23, 2008 in the Indiana University Department of Informatics Design House at 611 Woodlawn Avenue. We scheduled three successive user sessions, allowing enough time between users to sufficiently prepare our testing environment for the next session. The other users were not present in the room as we were conducting each individual test.

SESSION LENGTH

Our testing sessions lasted forty-five minutes on average, including introductions, consent forms, the pre-test questionnaire, actually conducting the usability tasks and scenarios, and the post-test feedback questionnaire. Approximately 20 minutes were dedicated to directly interacting with the Many Eyes website and carrying out the tasks and scenarios themselves.

TEST SETUP

We conducted our usability tests in the first floor "living room" of the Informatics Design House. The environment is similar to that of a typical student home in Bloomington.

- **Hardware:** Our users used a late-2008 MacBook Pro laptop with OS X version 10.5.5, and a corded Logitech USB mouse to interact with the Many Eyes website. To avoid distractions, we created a new, blank user account on this computer solely for the purpose of conducting this test. We used the built-in video camera and microphone to record both video and audio for the duration of each usability test.
- **Software:** Each user accessed the Many Eyes website using the Mozilla Firefox version 3.0.4 web browser, with both Javascript and Java enabled. We used Silverback² version 1.0.1, a streamlined usability testing application, for recording the user's activity. Silverback allowed us to simultaneously record audio and video, as well as all screen and mouse activity, while still being transparent and non-obtrusive for the user.
- **Session Recording Plan:** In addition to recording the screen, audio and video for each usability test, one researcher was responsible for filling out the Time Data Logging Form (which included observation notes), and a second researcher recorded all observations using the Data Logging Form. While the facilitator was next to the user for the duration of the usability test, these researchers were seated behind the user.

DESCRIPTION AND RATIONALES FOR THE TESTING SCENARIOS

Two scenarios were designed with the target audience in mind. The first scenario was "As an assignment for a business class, you must prepare a report on companies by country. You decide to include a visualization in your report." The second scenario was "You want to visualize President-elect Obama's references to race in his "A More Perfect Union" speech." Both scenarios were described in terms of end goals. The first scenario required the use of an existing numeric dataset, while the second was text-based. The second scenario was more difficult in that the user needed to register and login to Many Eyes in order to upload the text. The two scenarios covered everything specified in our test objectives.

DESCRIPTION AND RATIONALES OF THE TASKS

The tasks in the first scenario were sequenced to ease the user into experiencing the site by first using the search feature to find a particular dataset and viewing visualizations created by other users. Following that, the user was asked to use the dataset to create a specific type of visualization, modify the parameters of the visualization, publish it, and save a copy locally. The tasks in the second scenario were generally easier, but the user first had to register and login to accomplish them. We knew from our own experience with the site that the users might find the registration process problematic, and we wanted to see if they would persist and figure out how to accomplish those tasks. In all, the tasks covered the main functions and features of the Many Eyes site.

EVALUATION METHODS AND RESULTS

USABILITY GOALS

If we were actually working for Many Eyes (ideally during the designing and prototyping of the site), we would meet with key stakeholders to understand their perspectives regarding the important features of the site and their goals for the user's experience. We would then prioritize these features/goals, translate them into research questions, and formulate our scenarios and tasks to answer those questions. We chose to focus on ease of learning, and we selected features that we thought would be most useful to our target audience.

SUMMARY OF EVALUATION STRATEGIES

We used a pre-test questionnaire to gather information about the participants' backgrounds and familiarity with computers, the Internet, and data visualization software; we used a post-test questionnaire to solicit their thoughts and feelings immediately after using the Many Eyes site. The participants were asked to think aloud while working on the tasks, while observers documented the time spent on each task, the strategies and actions used to complete the tasks, and the users' comments. For the purposes of comparison, we trained a person to use the Many Eyes site by showing her the main features and letting her practice using the site for 30 minutes. At the end of that training, we had her complete the scenarios/tasks while being timed.

DESCRIPTION OF DATA ANALYSIS METHODS AND SELECTION RATIONALES

We entered time log data into Excel, calculated averages for each task, and entered the trained user's times for comparison. We think that the ability/time to complete tasks is an important metric for this site and for our target population, in particular because they generally have little spare time to spend learning new software.

We combined the observer's comments for all users by task with those of the time loggers (who made less-detailed notes than the observer) and then replayed the videos/screen captures for

clarification. We then summarized the comments by task as usability issues. If we were doing a full usability report, we would have examined the recordings more closely to find problems that were missed or were not noted during the test.

We tallied the results of the pre-test questionnaire so that we could describe our participants in our report. We combined and analyzed their responses on the post-test questionnaire. We then compared the responses with our summary of observations to confirm that our perceptions of problems matched their responses. If there had been inconsistencies, we would have reviewed the recordings for clarification.

SUMMARY OF DATA ANALYSIS

QUANTITATIVE DATA ANALYSIS

Our quantitative data include tallies of responses to a pre-test questionnaire (demographic and behavioral data) and measures of task completion and task time.

- *Number of tasks successfully completed*
All users completed all tasks with one exception: One user was unable to complete task 1.2 (View existing visualizations of the dataset).
- *Time spent on each task*
Table 1 shows the amount of time each user spent on each task, along with the average for the three users. For comparison purposes, we included the times for an experienced user. Figure 1 shows a chart of these data by user and task, and Figure 2 also shows cumulative time.

Table 1

Scenario.Task	User 1	User 2	User 3	Average	Exp. User
1.1	0:02:30	0:01:15	0:01:23	0:01:43	0:00:13
1.2	0:03:25	0:03:10	0:02:30	0:03:02	0:00:15
1.3	0:01:10	0:01:35	0:01:05	0:01:17	0:00:20
1.4	0:01:25	0:01:50	0:01:04	0:01:26	0:00:30
1.5	0:00:15	0:00:40	0:01:09	0:00:41	0:00:15
2.1	0:04:10	0:06:55	0:03:15	0:04:47	0:01:55
2.2	0:04:15	0:01:25	0:03:22	0:03:01	0:00:40
2.3	0:00:10	0:00:20	0:00:50	0:00:27	0:00:15
2.4	0:00:15	0:00:05	0:01:48	0:00:43	0:00:05
2.5	0:00:25	0:00:55	0:00:31	0:00:37	0:00:10
2.6	0:00:30	0:01:00	0:00:40	0:00:43	0:00:20

Figure 1: Time per task

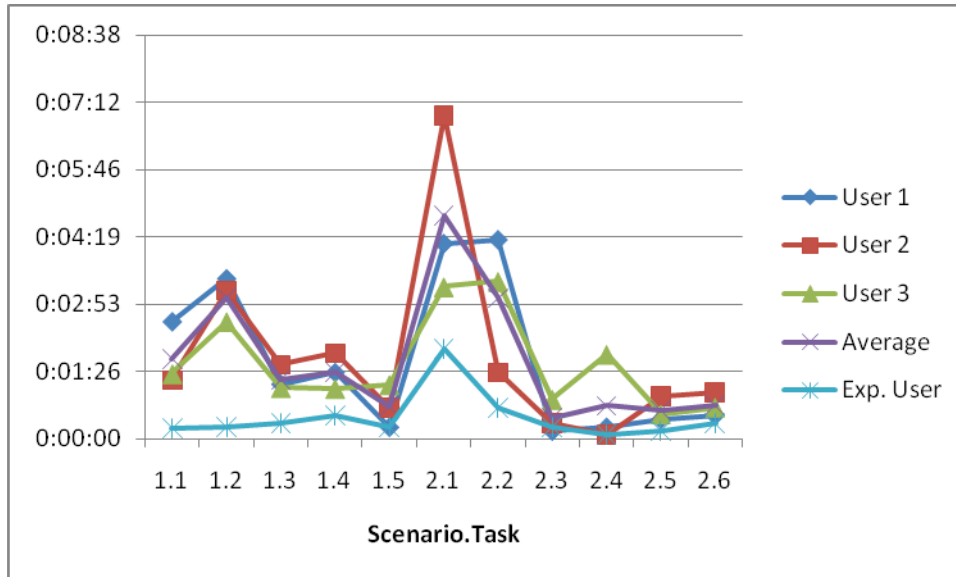
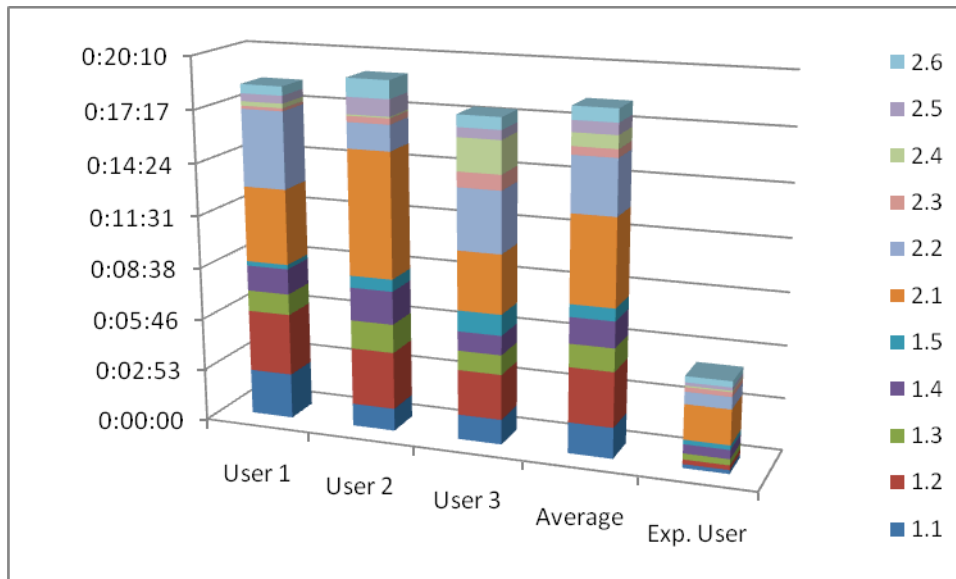


Figure 2: Cumulative time per user



If we were doing a full usability report, we would discuss these data in detail to prepare the reader for our reporting of usability problems and recommendations. In brief, Carolyn Snyder (as cited in Kuviaskey, 2003) suggests that a novice user will take anywhere from 3 to 10 times longer than an experienced user to complete a task.³ The performance data indicate problems related to Tasks 1.1 (Find the dataset for Forbes Global 2000 companies by country), 1.2 (View existing visualizations of the dataset), 2.1 (Register with the site and login), and 2.2 (Upload the text of the speech). The users' average time for all tasks was 3.71 times longer than the experienced user's time. For these four problematic tasks, the ratios were 7.90, 12.11, 2.49, and 4.52 respectively. While the ratio for Task 2.1 (2.49) is not extreme, the task itself takes longer

than any other task and is therefore a potential source of frustration for a busy user. Our analysis of the observation data and post-test questionnaire data confirmed this. Note that User 3 took significantly longer to complete Task 2.4, but that was because she chose to explore some additional features before completing the task.

QUALITATIVE DATA ANALYSIS

Our qualitative data include notes based on observation of a think-aloud protocol and responses to open-ended questions on a post-test questionnaire. If we were doing a full usability report, we would include our summaries of observations along with representative quotations from users. We would compare our observation data with post-test questionnaire responses. In brief, two users found Task 2.1 to be the most difficult (the third user indicated that almost everything, including Task 2.1, was difficult), and all found it to be the most “annoying or frustrating” aspect of the site. Two users especially enjoyed visualizing text as a Wordle.

CLASSIFICATION OF THE SEVERITY OF THE PROBLEMS AND RATIONALES

If a task takes too long to learn or to complete (or cannot be completed), users are likely to seek a faster approach (e.g. another tool, a different website). In a full usability report, we would classify the severity of the problems that we identified based on both our quantitative and qualitative data as follows:

PROBLEM SEVERITY

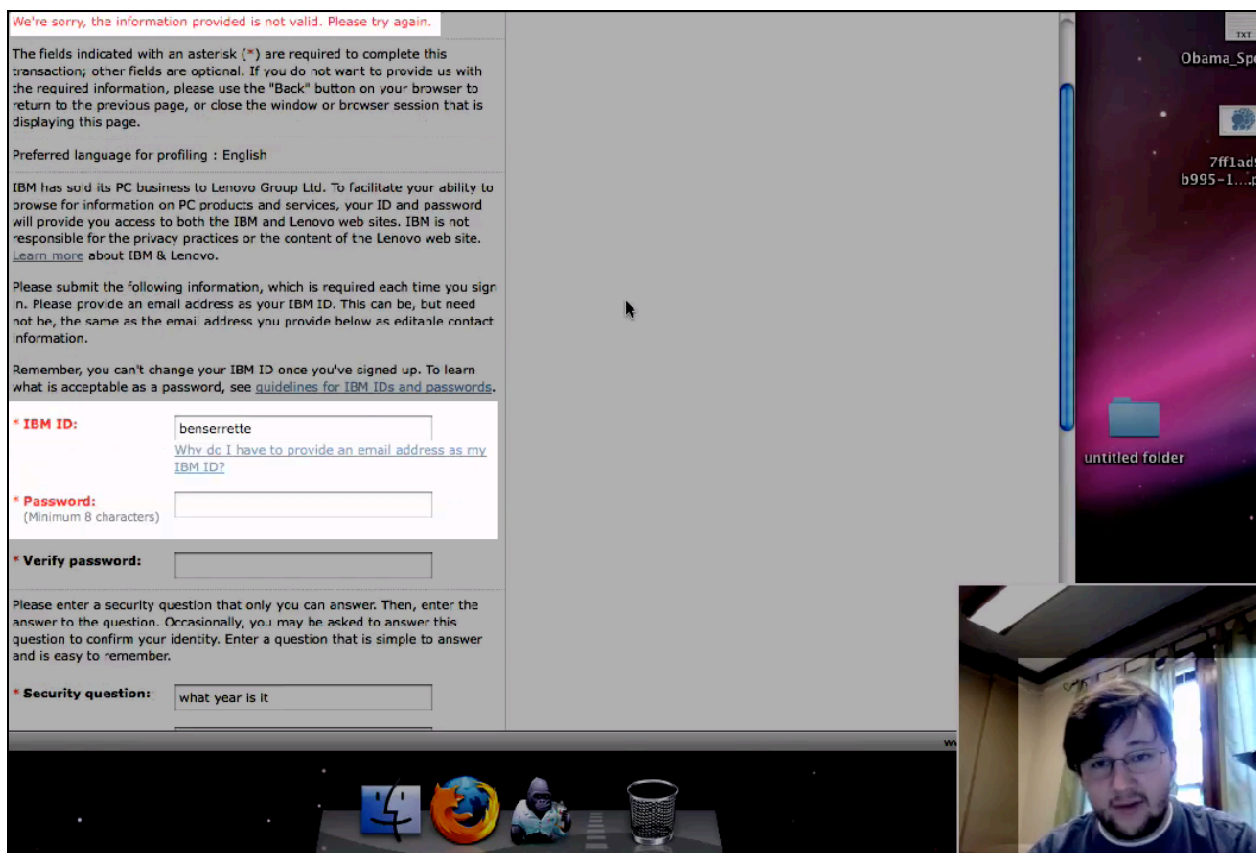
- 0 = Lack of success: One or more users were unable to complete the task.
- 1 = Significant delay: The task took so long that the user expressed frustration, or the task took longer than expected when compared with an experienced user’s time.
- 2 = Slight delay: The time to complete the task did not elicit frustration and was reasonable when compared with an experienced user’s time.
- 3 = No delay: The time to complete the task was close to an experienced user’s time.

Using this schema, tasks with the lowest numbers would have the most severe problems, while tasks with the highest numbers would not likely be cause for concern.

SCREENSHOTS FROM THE TESTING TO DOCUMENT FINDINGS

If we were doing a full usability report, we would use screenshots from the usability testing to illustrate the most severe problem areas and the user reactions. As an example, below is an illustration of a problem associated with Task 2.1 (Figure 3). All users had problems registering with the site, and the nature of those problems was not clarified by the feedback they received. In this example, relevant areas have been highlighted. Note the distance from the error message to the problem area. The error message says “We’re sorry, the information provided is not valid. Please try again.” The actual problem is that the “IBM ID” field requires an e-mail address, but all users initially entered text strings which conformed to normal user IDs. When the page returned the error message, the password that the user had previously entered was removed, leading some users to conclude that the problem was with the password.

Figure 3: Task 2.1 (Registration)



As another example, below is a screenshot with callouts that illustrates why users had difficulty with Task 1.2 (Figure 4).

Figure 4: Task 1.2 (Existing visualizations)

The screenshot shows the 'many eyes' website interface. The main content area displays a dataset titled 'Forbes Global 2000 companies by country'. The data is presented in a table with the following columns: Country, Count of Company, Sum of Sales (\$M), Sum of Profits (\$M), Sum of Assets (\$M), and Sum of Market Value (\$M). The table lists 14 countries, with Australia having the highest count at 50.

Country	Count of Company	Sum of Sales (\$M)	Sum of Profits (\$M)	Sum of Assets (\$M)	Sum of Market Value (\$M)
1 Australia	50	386.4	62.9	2421.7	855.4
2 Austria	13	103.7	9.3	562.7	153.9
3 Bahrain	2	2.6	0.4	43.2	7.2
4 Belgium	12	282.0	16.7	1626.9	191.4
5 Bermuda	24	206.4	17.2	489.6	241.9
6 Brazil	34	344.5	43.4	1122.3	795.2
7 Canada	59	701.6	76.7	3634.3	1111.2
8 Cayman Islands	4	24.8	6.7	64.1	81.9
9 Jersey	1	0.2	0.3	9.1	3.8
10 Chile	7	31.9	2.4	80.6	48.7
11 China	70	629.3	73.3	3864.6	2937.1
12 Colombia	2	4.8	0.6	37.3	10.7
13 Czech Republic	1	7.7	1.3	17.6	43.9
14 Denmark	9	119.3	11.0	628.9	165.9

On the right side of the screenshot, a yellow callout box contains the following text:

Task 1.2
This navigation element is below the bottom of the screen and separate from other navigation elements. Users did not see it when looking for existing visualizations of this dataset.

Below the callout box, another callout box points to a 'Visualizations of this data set' section, which includes a row of visualization icons and the text 'Part of these topic hubs' and 'Being watched by'.

USABILITY PROBLEMS AND RECOMMENDATIONS

If we were doing a full usability report, we would list all of the problems that we found in order of importance, describe why users had these problems, and make design recommendations for correcting the problems and improving usability. Task 2.1 (Registration) took the most time and was clearly the greatest cause of frustration for users. Task 1.2 (Existing visualizations) took much longer than expected, and one user gave up before finishing the task. As an example, for Task 1.2 we would describe the nature of the problem and recommend a possible solution. Figure 5 illustrates a recommended solution in which the navigation for “Visualizations of this data set” is brought above the bottom of the screen and combined with other navigation elements. Most users clicked the “Visualize” button when searching for existing visualizations; by juxtaposing these two elements, users are more likely to recognize the difference and choose the appropriate one.

Figure 5: Task 1.2 (Existing visualizations) recommendation

Task 1.2
Combine these navigation elements to avoid confusion and improve usability.

Visualizations of this data set

watch this | add to topic hub | visualize | rate this

watch this | add to topic hub | visualize | rate this

Visualizations of this data set

- Part of these topic hubs
- Being watched by

Country?	Count of Company	Sum of Sales (\$M)	Sum of Profits (\$M)	Sum of Assets (\$M)	Sum of Market Value (\$B)
1 Australia	50	386.4	62.9	2421.7	856.4
2 Austria	53	103.7	9.3	462.7	163.9
3 Bahrain	2	2.6	0.4	43.2	7.2
4 Belgium	12	282.0	16.7	1626.9	191.4
5 Bermuda	24	206.4	17.2	469.6	241.9
6 Brazil	34	344.5	43.4	1122.3	795.2
7 Canada	59	701.6	76.7	3634.3	1111.2
8 Cayman Islands	4	24.8	6.7	54.1	81.9
9 Jersey	1	0.2	0.3	9.1	3.8
10 Chile	7	31.9	2.4	80.6	48.7
11 China	70	629.3	73.3	3864.6	2037.1
12 Colombia	2	4.8	0.6	37.3	10.7
13 Czech Republic	1	7.7	1.3	17.6	43.9
14 Denmark	9	119.3	11.0	628.9	166.6

Based on the feedback we received on our usability testing preparation packet, we made the following revisions:

- We modified the Consent Form to indicate the length of the testing session.
- We removed scenario 2 and modified scenario 3 so that the registration and login process was first. We also added written instructions so that the participant could read them in addition to having the moderator introduce them.
- We modified the Time Logging Form to optimize the use of the page.
- During the post-test questionnaire, we made sure that the participants had a copy of the scenarios and tasks for reference.

We have included our entire revised usability testing preparation packet in a separate compressed file for reference.

REFERENCES

- ¹ Many Eyes. Retrieved November 28, 2008 from <http://manyeyes.alphaworks.ibm.com/manyeyes/page/About.html>.
- ² Clearleft Ltd. (2008). Silverback [Software]. Available from <http://silverbackapp.com/>.
- ³ Kuvias, M. (2003). *Observing the user experience: A practitioner's guide to user research*. San Francisco: Morgan Kaufmann Publishers.