

1. Report of data collected, including both quantitative and non-quantitative observations

1.1 The experiment

We ran our experiment with 5 test subjects.

Each subject performed the list of tasks (shown below) on two different interfaces, “pull down menu” and “pop up new page”, in random order. And we collect the quantitative data and non-quantitative observations during/after the experiment sessions.

The variables we recorded during/after the experiment are as follows:

- 1) The time the subject took to perform each task using different interfaces.
- 2) The errors the subject made during the test. We will use this data to calculate Error Rate after collecting the data.
- 3) The satisfaction rates the subject gave after performing test on each interface.
- 4) The time the subject spent to learn to use it to proficiency. (The test subject will let us know when he/she believes he/she has learnt to use this interface proficiently.)

The task list:

- 1) View detailed information of class “Human Computer Interactn”
- 2) Go back to class list.
- 3) View status of class “Web Engineering”
- 4) View grade of class “Intro Machine Learning”

1.2 The data collected

The data we collected are as follows:

1.2.1 quantitative data

1. “Pull down menu”:

		subject1	subject2	subject3	subject4	subject5
Time to perform each task (s)	task1	0.93	1.16	1.01	1.03	0.98
	task2	7.73	1.34	0.66	0.98	1.85
	task3	2.14	1.22	0.95	0.85	1.02
	task4	1.07	0.95	1.11	1.13	1.16
Number of errors made (e.g. clicked the wrong button)	task1	0	0	0	0	0
	task2	2	0	0	0	0
	task3	0	0	0	0	0
	task4	0	0	0	0	0
Satisfaction rate (0 not satisfied - 10 very satisfied)		8	10	9	9	9
Time spent to learn to use the interface proficiently. (s)		9.83	4.09	3.83	5.78	4.67

Table 1.1 Quantitative data for “pull down menu” interface

2. "Pop up new page"

		subject1	subject2	subject3	subject4	subject5
Time to perform each task (s)	task1	0.88	1.05	0.85	0.95	0.87
	task2	0.54	0.92	0.91	0.58	0.66
	task3	1.70	1.98	0.68	0.60	0.71
	task4	1.11	1.75	1.18	1.15	1.24
Number of errors made (e.g. clicked the wrong button)	task1	0	0	0	0	0
	task2	0	0	0	0	0
	task3	0	0	0	0	0
	task4	0	0	0	0	0
Satisfaction rate (0 not satisfied - 10 very satisfied)		10	9	8	10	10
Time spent to learn to use the interface proficiently. (s)		2.06	3.98	1.05	2.13	1.43

Table 1.2 Quantitative data for "pop up new page" interface

1.2.2 non-quantitative observations

- 1) Subject 1: When testing "Pull down menu" interface, subject 1 had trouble in looking for the button to go back to class list. He finally figured it out after clicking the wrong button 2 times: clicking home page button and clicking outside of the class name area.
- 2) Subject 2: Subject 2 had no trouble in completing all tasks. The whole process went smoothly.
- 3) Subject 3: Subject 3 had no trouble in completing all tasks. The whole process went smoothly. Besides, she said "great" multiple times when using pull down version.
- 4) Subject 4: Subject 4 had no trouble in completing all tasks. The whole process went smoothly.
- 5) Subject 5: Although subject 5 made no error in completing tasks, he said "uh" which illustrated her hesitation and confusing in completing task2 in pop up version.

2. Report of data analysis and summary

2.1 Statistic summarizing

		“Pull down menu”		“Pop up new page”	
		Mean	Standard Deviation	Mean	Standard Deviation
Time to perform each task (s)	task1	1.022	0.08585	0.92	0.0815
	task2	2.512	2.95027	0.722	0.18144
	task3	1.236	0.52319	1.134	0.65328
	task4	1.084	0.08173	1.238	0.26269
Number of errors made (e.g. clicked the wrong button)	task1	0	0	0	0
	task2	0.4	0.89443	0	0
	task3	0	0	0	0
	task4	0	0	0	0
Satisfaction rate (0 not satisfied - 10 very satisfied)		9	0.70711	9.4	0.89443
Time spent to learn to use the interface proficiently (s)		5.64	2.45943	2.13	1.12736

Table 2.1 Statistic summary (mean and standard deviation).

2.2 Graphing techniques

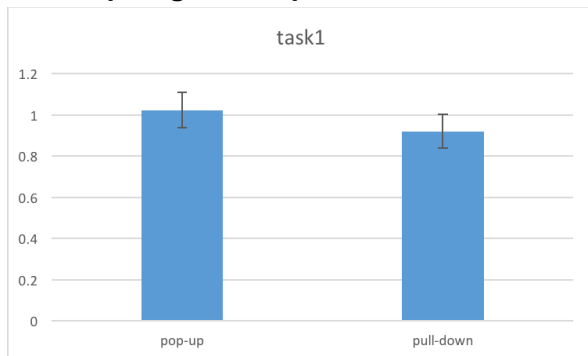


Figure 2.1 Standard error of the mean of task1 time

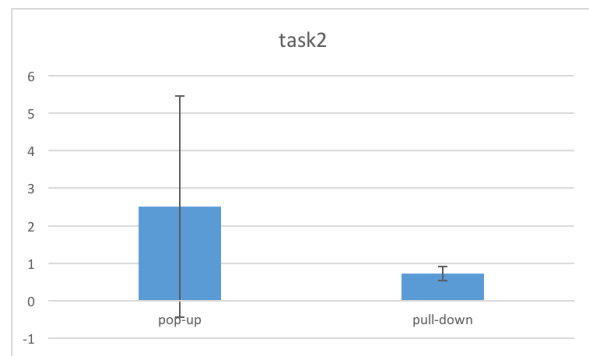


Figure 2.2 Standard error of the mean of task2 time

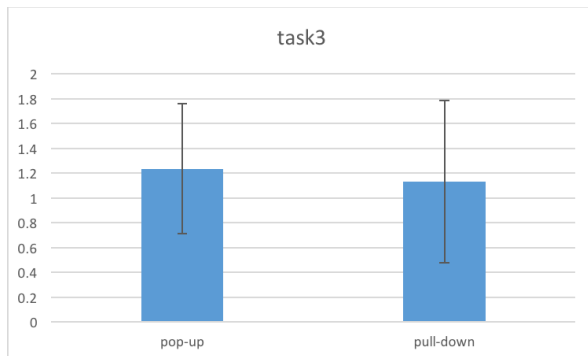


Figure 2.3 Standard error of the mean of task3 time

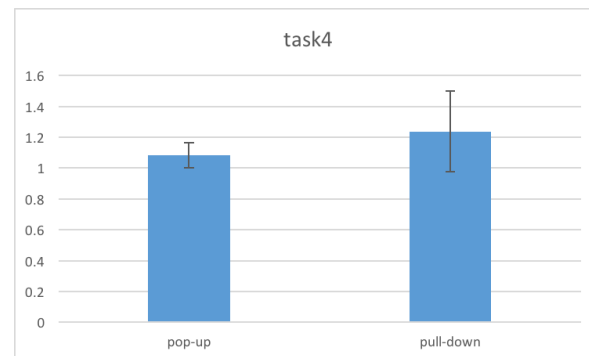


Figure 2.4 Standard error of the mean of task4 time

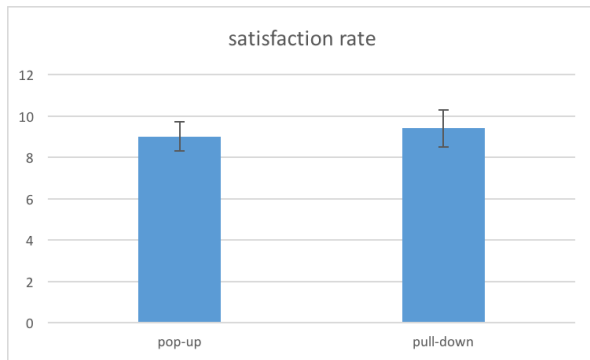


Figure 2.5 Standard error of the mean of satisfaction rate

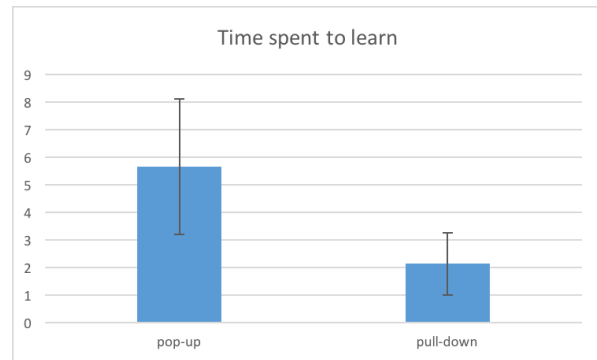


Figure 2.6 Standard error of the mean of time in learning

2.3 t tests

Note:

* All calculations were done based on 0.05 significance level. (95% confidence interval)

** We would use a two-tailed test to see if two means are different from each other. While in one-tailed test, we tested whether one mean was higher than the other. We were not interested in whether the first mean was lower than the other, only if it was higher. So we were only interested in two-tailed test.

1. Task1 time:

Task1 Time	"Pull down menu"	"Pop up new page"
Mean	1.022	0.92
Variance	0.01	0.01
Observations	5	5
Pooled Variance	0.01	
Hypothesized Mean Difference	0	
df	8	
t Stat	1.923	
P(T<=T) one-tail	0.045	
t Critical one-tail	1.860	
P(T<=T) two-tail	0.091	
t Critical two-tail	2.306	

Table 2.2 t test results of task1 time.

Conclusion: from two-tailed test, the t-value is 1.923 and the p-value is 0.091. The result is not significant at $p < 0.05$. So this difference is considered to be not quite statistically significant.

2. Task2 time:

Task2 Time	"Pull down menu"	"Pop up new page"
Mean	2.51	0.72
Variance	8.7	0.03
Observations	5	5
Pooled Variance	4.37	
Hypothesized Mean Difference	0	
df	8	
t Stat	1.354	
P(T<=T) one-tail	0.107	
t Critical one-tail	1.860	
P(T<=T) two-tail	0.214	
t Critical two-tail	2.306	

Table 2.3 t test results of task2 time.

Conclusion: from two-tailed test, the t-value is 1.354 and the p-value is 0.214. The result is not significant at $p < 0.05$. So this difference is considered to be not quite statistically significant.

3. Task3 time:

Task3 Time	"Pull down menu"	"Pop up new page"
Mean	1.24	1.13
Variance	0.27	0.43
Observations	5	5
Pooled Variance	0.35	
Hypothesized Mean Difference	0	
df	8	
t Stat	0.273	
P(T<=T) one-tail	0.396	
t Critical one-tail	1.860	
P(T<=T) two-tail	0.792	
t Critical two-tail	2.306	

Table 2.4 t test results of task3 time.

Conclusion: from two-tailed test, the t-value is 0.273 and the p-value is 0.792. The result is not significant at $p < 0.05$. So this difference is considered to be not quite statistically significant.

4. Task4 time:

Task4 Time	"Pull down menu"	"Pop up new page"
Mean	1.08	1.29
Variance	0.01	0.07
Observations	5	5
Pooled Variance	0.04	
Hypothesized Mean Difference	0	
df	8	
t Stat	-1.636	
P(T<=T) one-tail	0.070	
t Critical one-tail	1.860	
P(T<=T) two-tail	0.140	
t Critical two-tail	2.306	

Table 2.5 t test results of task4 time.

Conclusion: from two-tailed test, the t-value is -1.636 and the p-value is 0.140. The result is not significant at $p < 0.05$. So this difference is considered to be not quite statistically significant.

5. Satisfaction Rate:

Satisfaction Rate	"Pull down menu"	"Pop up new page"
Mean	9.0	9.4
Variance	0.5	0.8
Observations	5	5
Pooled Variance	0.65	
Hypothesized Mean Difference	0	
df	8	
t Stat	-0.784	
P(T<=T) one-tail	0.228	
t Critical one-tail	1.860	
P(T<=T) two-tail	0.455	
t Critical two-tail	2.306	

Table 2.6 t test results of satisfaction rate.

Conclusion: from two-tailed test, the t-value is -0.784 and the p-value is 0.455. The result is not significant at $p < 0.05$. So this difference is considered to be not quite statistically significant.

6. Learning:

Time to be proficient	"Pull down menu"	"Pop up new page"
Mean	5.64	2.13
Variance	6.05	1.27
Observations	5	5
Pooled Variance	3.66	
Hypothesized Mean Difference	0	
df	8	
t Stat	2.901	
P(T<=T) one-tail	0.010	
t Critical one-tail	1.860	
P(T<=T) two-tail	0.020	
t Critical two-tail	2.306	

Table 2.6 t test results of learning.

Conclusion: from two-tailed test, the t-value is 2.901 and the p-value is 0.020. The result is significant at $p < 0.05$. So this difference is considered to be quite statistically significant.

3. Discussion of the result

3.1 Discussion about the mean and standard deviation

1) By comparing the mean of speed and standard deviation for time spent on task1 to task4, we find that the average speed for "pop up" is faster than the average speed for "pull down". And the standard deviation of completing task2 using pop up is the biggest, which indicates that the highest difficulty of this task (It makes sense because task2 is the first time asking user to go back to the classes list page. As a result, users don't have any prior knowledge of completing it. Some users don't know where to click in pop up version. While in the pull down version, there is a back button showing in the page, which clearly gives user instructions of the operation. Clicking class name again to hide class information in pop up is not that intuitive to every user).

2) By comparing the mean of speed and standard deviation for learning to use the system proficiently, we find that the average speed for learning to use "pop up" version is faster than learning to use "pull down" version. In the experiment, we noticed that users seem to consider they've learned how to use "pop up" version after completing task1 and task2. While in using "pop up" version, users seem to get familiar with the operations after completing all tasks and some of users even need to try extra operations afterwards to proficiently use. This difference came from the lack of instructions in pull down version.

3) By comparing the mean and standard deviation for satisfaction rate, we find that 3/5 users prefers pull down while 2/5 users prefer pop up. However, the difference between satisfaction rate of pop up and pull down of the same user is quite small and satisfaction rate for these two versions are both at the high level, which indicates that these two versions are both easy to use while different users might have different preference.

4) By comparing the mean and standard deviation for number of errors in completing each task, we can easily draw conclusion from the results that nearly no errors would occur when using both version except in task2 which also results from the lack of instructions in pull down version design.

3.2 Discussion about the t-test

We would used a two-tailed test to see if the means in task finish time, satisfaction rate and learning time of the two interfaces are different from each other.

From the results, we can get the following conclusion that:

1) For all 4 tasks, the difference between the means of the task finish time **is not** statistically significant (since the t value is smaller than the critical value). So although the average speed for “pop up” is faster than the average speed for “poll down” (as we shown above in statistic summary part), the differences are not considered to be statistically significant.

2) For satisfaction rate, the difference between the two means of the satisfaction rate **is not** statistically significant (since the t value is smaller than the critical value).

3) For learning time, the difference between the two means of the learning time **is** statistically significant (since the t value is larger than the critical value). So users took less time in average to learn “pop up” interface than “pull down” interface and the difference is considered to be statistically significant.

Thus, the two interfaces show statistically significant difference in learning time, but in other aspects, there is no notable difference observed.

wn” is easier and faster to use.