

Final Iteration – Quantitative Evaluation

Team TuTasty / Moderated, Recorded and Analysed by Xinyi Miao, Maja Schuknecht

Quantitative analyses were conducted on 10 test subjects aged 24 to 30 years. The specific procedures for data collection and Hypothesis, Test cases, and Metrics have been described in detail in the report "Preparation" of Iteration 4.

Demographic Data

A total of 10 people took part in the user test. Four males and five females and 1 diverse were obtained. The age of the test participants ranged from 24 to 30 years (MW=25.8). Most of the subjects were students and 60% of them went to the cafeteria more than 2-4 times per week.

Gender
10 responses

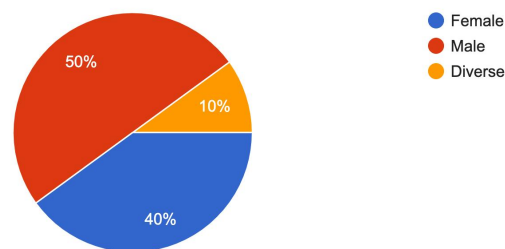


Figure 1: Gender distribution

Occupation
10 responses

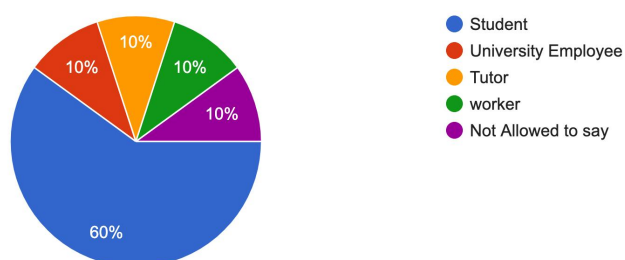


Figure 2: Occupation distribution

How often do you eat at the mensa?
10 responses

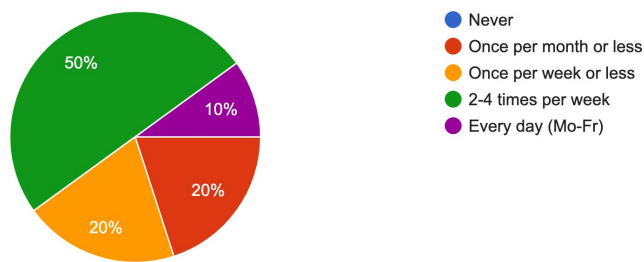


Figure 3: Frequency distribution

Results of the hypothesis

These hypothesis are in line with our functional and usability requirements. Subjects were asked to perform a total of 10 tasks. We observed and recorded the required data, and based on the metrics, the following results were calculated.

Goals	Hypothesis	Objective measures	Result
1. Learnability	1.1 More than one way is used to cancel dishes.	The number of different commands to fulfill a specific task.	2 ways
2. Effectiveness	2.1 Average completion rate of >70% for all tasks completely performed by users	$\frac{\text{number of task which completed successfully}}{10} \times 100\%$	100%
	2.2 At least 80% of the test persons complete the task of booking dishes.	$\frac{\text{number of people who completed the task}}{\text{number of people who did not complete}} \times 100\%$	100%

	<p>2.3 At least 80% of the test persons complete the task of checking the location of the counter where the dish is located and find out the queue at this counter.</p>	$\frac{\text{number of people who completed the task 5}}{\text{total number of test persons}} \times 100\%$	100%
3. Efficiency	<p>3.1 The average time for a user to book a dish for the first time is less than 20s.</p>	The time users take to complete task 8.	2.4s
	<p>3.2 The average time to cancel a scheduled dish does not exceed 20s.</p>	The time users take to complete task 9.	1.7s
4. Memorability	<p>4.1 At least 70% of users can remember the three first-level navigation of the product after testing.</p>	$\frac{\text{number of people who remembered navigation}}{\text{total number of test persons}} \times 100\%$	60%
5. Errors	<p>5.1 The average number of errors does not exceed 6 in fulfilling all tasks. (wrong tasks executed, wrong clicks, too many clicks)</p>	The number of errors for all tasks.	2.6
	<p>5.2 At most 40% of the test persons will be confused or make mistakes with the counter map.</p>	$\frac{\text{number of people who make error by task 4}}{\text{total number of test persons}} \times 100\%$	70%
6. Subjective	<p>6.1 The prototype achieves at least a</p>	Rating number	8.8

satisfaction	rating of 6.5 out of 10 on average.		
	6.2 SUS-Score exceeds a value of 70.	SUS-Score	89.5
	6.3 The average Task Level Satisfaction does exceed a value of 5 ⁽⁵⁾ .	SEQ-Score	6.52

Table 1 Results of hypothesis testing

Hypothesis	Approved or rejected
1.1	Accepted
2.1	Accepted
2.2	Accepted
2.3	Accepted
3.1	Accepted
3.2	Accepted
4.1	Rejected
5.1	Accepted
5.2	Rejected
6.1	Accepted
6.2	Accepted
6.3	Accepted

Table 2 Overview over approved or rejected hypothesis

Based on the results in the above table, we can see that hypothesis 4.1 and hypothesis 5.2 are rejected and all other hypotheses are proved. That is, 4 of

the 6 objectives are satisfied, while "Memorability" and "Errors" are still not satisfied.

For hypothesis 4.1, 60% of subjects remember the three first level navigation after testing, the others remember only two first level navigation .

For hypothesis 5.2, 70% of subjects made the wrong click on the task of viewing the counter location.

They think the counter number is clickable and thus filter what dishes are available at each counter. And in fact, right now we don't have this feature, but we will consider to continue developing this feature in the future.

For subjective measure, The average score of SUS is 89.5, Over 85, so our prototype meets the “excellent” rating^[1]. And result of the Overall Rating is 8.8, are similar to SUS. SEQ-Score is 6.52, subjects generally find most of the tasks are very easy.

Taking hypothesis 6.2 System Usability Scale (SUS) as an example, the following is the specific hypothesis testing process.

Descriptive Statistics

	Valid	Missing	Mean	Std.Deviation	Min.	Max.
SUS Score	10	0	89.5	7.976	77.5	97.5

Figure 4: Descriptive statistics of the System Usability Scale (SUS)

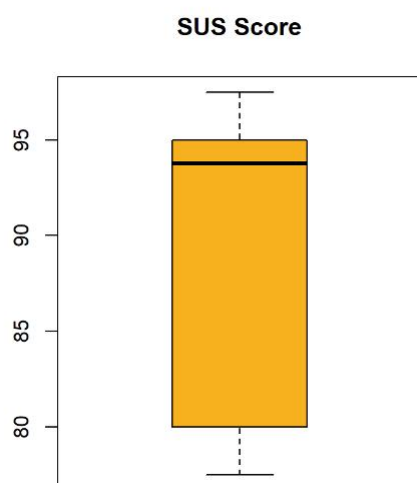


Figure 5: Boxplot of the System Usability Scale (SUS)

H0: Subjects do not exceed a System Usability Scale of 70 regarding the complete application.

H1: Subjects do exceed a System Usability Scale of 70 regarding the complete application.

→ $M = 89.5$, $SD = 7.976$

→ H0 can be rejected.

The system usability score reached 89.5. This value reflects an acceptability rating of "Acceptable" and an adjective rating of "Excellent"^[1]. Thus, this value exceeds the minimum value of 70 expected in the hypothesis, which means that the application has no critical usability defects.

Discussion

TUtastyApp is considered to be very intuitive, which is also confirmed by the excellent system usability scale. The particularly attractive design was repeatedly praised. The general concept of the software was also rated as very helpful and useful.

In summary, it can be said that the hypothesis concerning learnability could be accepted, which suggests that processes within the app can be easily learned.

All three hypotheses regarding goal 2 - Effectiveness, resulted in 100%, which means that TUtastyApp achieves the goal of effectiveness perfectly and the testee can easily complete all tasks, which corresponds to the SEQ result of 6.52. Therefore, in general, TUtastyApp is easy to use^[2].

The execution of certain tasks described in hypotheses 3.1-3.2 could also be completed by the test subjects in a desirable time frame on average. This underlines the efficiency of the system.

It is important to note that overall, the application still received a desirable average rating of 8.8 out of 10, with a sus result of 89.5, which shows that TUtastyApp has excellent usability and user satisfaction^[3].

However, several implications for future design and functional changes to the application can be drawn from the reported results. Special attention should be

paid to the application's performance on the feature of counter maps. For the future development of the application, for example, consider adding the feature of dish favorites. As well as for the optimization of the rating page. One point worth noting is the problem of "queue length is occluded by highlight", which can be solved by highlighting only the numeric part. The prototype of the application will be further improved in this regard in the next iteration step.

References

- [1] Aaron Bangor, Philip T. Kortum & James T. Miller (2008) An Empirical Evaluation of the System Usability Scale, International Journal of Human–Computer Interaction, 24:6, 574-594, DOI: [10.1080/10447310802205776](https://doi.org/10.1080/10447310802205776)
- [2] "10 Things To Know About The Single Ease Question (SEQ)", <https://measuringu.com/seq10/>, Retrieved on 12.01.2023
- [3] "System Usability Scale (SUS)", <https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html>, Retrieved on 12.01.2023
- [4] "Usability Metrics – A Guide To Quantify The Usability Of Any System", <https://usabilitygeek.com/usability-metrics-a-guide-to-quantify-system-usability/>, access on 09.01.2023