CS348 User-Based Evaluation Report

Cem Yilmaz, Hawari Ansari, Damian Kwok, Viktor Saprykin

March 2024

Contents

1 Introduction

For this project, we were provided with a flawed prototype of an online shopping website, which contained a wide range of usability problems. These ranged from confusion-causing minor inconveniences that led users to take more time doing an action that needed (for example by not having a separation/filter of items by type causing users to scroll through the entire item list), to major accessibility issues responsible for outright frustration like not being able to close the cookies window obstructing the view or not being able to see the items in your basket.

This greatly limited the use of the site, and as such needed to be redesigned. Doing so was the focus of the first part of this coursework, in which many improvements and modifications were proposed and successfully implemented for both desktop and mobile devices.

The main goal of this part of the coursework is then to carry out a user-based usability evaluation of the redesigned website to test that:

1. *Our redesigns make the website more usable for desktop users, compared to the prototype.*
2. *The website provides equivalent usability for mobile users to the redesigned desktop interface.*

In order to conduct this user-based usability evaluation, we had to compare the different versions of the website. As such, we conducted the study with 9 participants in each category: prototype, redesigned desktop and redesigned mobile, totalling 27 participants for the entire study.

This choice, as well as our carefully chosen evaluation methodology, were picked in order to eliminate as much bias as possible. We will discuss the exact ways by which we achieved this in the following section. However, despite trying to minimise bias in participant selection, study environment and results, there were still a number of limitations which will additionally be discussed further below.

2 Evaluation design

2.1 Key Issues

To perform a proper user-based evaluation that would yield representative results many criteria have been taken into consideration. This is because there exists multiple methods to choose from, each varying in their respective approach to the problem and as such providing varying results from one another even given the same sample of participants.

We started by considering the standard key issues when it comes to conducting such a study. Specifically, the issues to think about were[1]:

- Setting goals:
  - How are we defining usability?
  - What data are we going to collect?
  - How are we going to collect it?
- Identifying participants:
  - Who are we gathering the data from?
  - Is our sample representative?
- Relationship with participants:
  - Do we provide an ability to give written consent?
- Triangulation:
  - Do we look at data from more than one perspective?
  - Do we collect more than one type of data?
- Pilot studies:
  - Do we do a small trial of main study?

2.2 Methodology

2.2.1 Setting goals

- **Defining Usability**

  For the purposes of this study we decided that usability will be defined as a combination of the quantitative data and the qualitative data, with quantitative being used as the primacy piece of data for statistical analysis while qualitative would be provide additional information and insights.

- **Data Collected**

  For the primary piece of quantitative data collected we decided to go with time taken for each participant to complete a list of tasks on the website. This aimed to measure the time-based efficiency[2] of the websites, which is an indicator of their general usability.

  For the primary piece of qualitative data collected we decided to go with a couple of measures for each participant, their satisfaction with using the website on a scale from one to five and the agreeableness extent to whether the website was usable/frustrating/confusing/intuitive. We also decided to collect some further qualitative data in the form of additional comments or observations that would be made by the interviewer during the study on the participant's strategy for the completion of the tasks. Additionally, we thought it would be a good idea to note any errors and their potential cause made by the participant during the study.

- **Collection Methods**

  There were a few method to choose from when it came to collecting the data. These were mainly lab studies, surveys and interviews. After some consideration we have decided to mainly focus on a combination of the lab study and a questionnaire.

  This approach consisted of observing a sample of users while they complete a list of tasks(given below) and measuring the time taken to complete the entire list in order to measure the usability of the websites. This would later on be followed by a questionnaire(also given below) who's purpose was to record the qualitative data. We decided that using the lab study as our main approach would make it possible to measure multiple variables that we considered important after hypothesising over the usability of the different website versions. Furthermore, conducting a lab study should also relate to the hypothesis or hypotheses we want to validate.

  Below is the exact task list and questionnaire that the participants got. There is also the testing script that was used by the group attached to the end of this report.

- **Task List**

  1. Add 3 Artichokes to the basket
  2. Add 3 fruits, 3 vegetables, and 3 other non-fruit/vegetable items to the basket
  3. Use the search bar to add 4 salmons, an orange juice, and a radish to the basket
  4. Add 5 pumpkins to the basket, go to checkout, and clear the basket afterward
  5. Add 2 breads to the basket, go to checkout, and add another bread
  6. Add 2 plums and 2 chocolate bars to the basket, go to checkout, and remove the chocolate bars
  7. Ensure the basket contains only 2 plums and 3 breads, then press "Pay" on the checkout page

- **Questionnaire**

  1. Rate your satisfaction with the website on a scale from 1 to 5

2. The website good usability wise. Choose between: Strongly Disagree, Disagree, Agree, Strongly Agree
3. I experienced frustration while using the website. Choose between: Strongly Disagree, Disagree, Agree, Strongly Agree
4. I experienced confusion while using the website. Choose between: Strongly Disagree, Disagree, Agree, Strongly Agree
5. Add 2 breads to the basket, go to checkout, and add another bread
6. Add 2 plums and 2 chocolate bars to the basket, go to checkout, and remove the chocolate bars
7. I found the website intuitive. Choose between: Strongly Disagree, Disagree, Agree, Strongly Agree

2.2.2 Identifying participants

As mentioned previously, a total of 27 participants took part in the study. Each participant was obtained from opportunity sampling and assigned to one of three groups Prototype, Redesigned Desktop, and Redesigned Mobile meaning nine participants were used for each group. The participants were distributed randomly, however, in order to eliminate as much bias as possible considerations were taken to ensure a roughly similar distribution of the internet proficiency of participants among groups. Further sample representation is discussed in section 4.

2.2.3 Relationship with participants

Before participating we made sure that every participant gives either a verbal or written consent, based on their choice, to take part in the study.

2.2.4 Triangulation

Multiple points of view are better than only one. This is what we had in mind when deciding on the combination of our main lab study with a questionnaire. It allowed us to easily collect and analyse more data. As it is easy and quick to deploy, the response time is pretty allowing us to note down user the users' satisfaction relatively quickly. The questions can be closed, in the form of multiple choices or ratings. This approach, although definitely provides less detail than open-ended questions, has a benefit of being a lot easier to analyse since the answers can be quantified. Open questions, on the other hand, allow the user to express more detailed thoughts, but are much more difficult to analyse, especially given a large sample size.

2.2.5 Pilot studies

As this study was relatively simple and required only 1 participant to be observed at the time we decided to proceed with the main study right away.

3 Results

3.1 Analysis of time taken

Our main goal in mind when redesigning the website was to make its use easier than the prototype given, and almost as easy on mobile platforms.

As this is a difficult metric to measure objectively, we opted to use the time taken to complete the list of task as a metric for ease of use. The reasoning behind this choice was our assumption that if tasks were completed faster, this would indicate that they experienced less difficulties performing the tasks, and as such would have used the site more effectively and efficiently.

The time required to complete the task by each group is displayed in the table below.

Group	Mean time(s)	stdev
Prototype	329.2	29.2
Redesigned Desktop	201.8	39.8
Redesigned Mobile	239.4	62.4

Mean and standard deviation of each group to complete the task

3.1.1 Prototype vs Redesigned Desktop

Using the the mean and standard deviation of both the Prototype and Redesigned Desktop from the table above, we perform a statistical analysis that would help us to test our first statement:

1. Our redesigns make the website more usable for desktop users, compared to the prototype.

We start by making the following hypotheses:

H0: Participants took the same amount of time to complete the task list on both the prototype and the redesigned website on desktop.

H1: Participants took less amount of time to complete the task list on the redesigned website on desktop than on the prototype.

Two-sampled t-test[3] was used to determine if the amount of time taken for each group were statistically different such that we could either accept or reject the null hypothesis.

Due to the fact that we only have two independent groups, the choice for which t-test to use; paired, two-sample or one-sample to analyse the results was obvious. Another way to confirm this would also be the use of an ANOVA test [4] to determine whether the variances of both groups were approximately equal. If the resulting chance would be higher than the significance level of say 0.05, we could assume that the variances are approximately equal and use a two-sample t-test.

First of all, we decided on the significance level and the degrees of freedom which would gave us the theoretical value from the t-distribution - critical t-value. For this test we assumed that the significance level is 0.005 as per standard practises and our degrees of freedom value was the sum of our sample sizes minus two so 9 + 9 - 2 = 16. Using the t-distribution table[5] we found that the critical t-value for these values was 1.746.

Then we calculated our test statistic - t with the following steps:

- Difference of group averages

$$329.2 - 201.8 = 127.4$$

- Pooled variance

$$v_p^2 = \frac{((n_1 - 1)s_1^2) + ((n_2 - 1)s_2^2)}{n_1 + n_2 - 2} = \frac{((9-1)29.2^2) + ((9-1)39.8^2)}{9 + 9 - 2} = 1614.35$$

- Pooled standard deviation

$$s_p^2 = \sqrt{v_p^2} = \sqrt{1614.35} = 40.18$$

- Test statistic

$$t = \frac{\text{Difference Of Group Averages}}{\text{Standard Error of Difference}} = \frac{127.4}{(40.18 \times \sqrt{\frac{1}{9} + \frac{1}{9}})} = 6.73$$

As the computed test statistic - t is larger than critical t-value, the results were statistically significant for us to reject the null hypothesis and thus allowed us to conclude that indeed participants took less amount of time to complete the task list on redesigned website on desktop than when using the prototype.

3.1.2 Redesigned Desktop vs Redesigned Mobile

Using the the mean and standard deviation of both the Redesigned Desktop and Redesigned Mobile from the table above, we perform statistical analysis that would help us to test our second statement:

2. The website provides equivalent usability for mobile users to the redesigned desktop interface.

We start by making the following hypotheses:

H0: Participants took the same amount of time to complete the task list on both the redesigned website on mobile and the redesigned website on desktop.

H1: Participants took more amount of time to complete the task list on the redesigned website on mobile than on the redesigned website on desktop.

Once again, we have used the two-sampled t-test with the the significance level being 0.005, the degrees of freedom being 16 and critical t-value being 1.746.

Then we calculated our test statistic - t with the following steps:

- Difference of group averages

$$239.8 - 201.8 = 38$$

- Pooled variance

$$v_p^2 = \frac{((n_1 - 1)s_1^2) + ((n_2 - 1)s_2^2)}{n_1 + n_2 - 2} = \frac{((9 - 1)39.8^2) + ((9 - 1)62.4^2)}{9 + 9 - 2} = 2738.9$$

- Pooled standard deviation

$$s_p^2 = \sqrt{v_p^2} = \sqrt{2738.9} = 52.3$$

- Test statistic

$$t = \frac{\text{Difference Of Group Averages}}{\text{Standard Error of Difference}} = \frac{38}{(52.3 \times \sqrt{\frac{1}{9} + \frac{1}{9}})} = 1.54$$

As the computed test statistic - t is smaller than critical t-value, the results were not statistically significant enough for us to reject the null hypothesis. This means that the participants took the same or less amount of time to complete the task list on a mobile device than on desktop and thus allowed us to conclude that indeed the redesigned website is as usable on mobile as it is on desktop.

3.2 Qualitative Results

3.2.1 Survey Results

Question	Prototype (Desktop)		Redesigned (Desktop)		Redesigned (Mobile)	
	Mean	Stdev	Mean	Stdev	Mean	Stdev
Rate your satisfaction using the website	1.9	0.9	4.2	0.6	4	0
The website is good usability wise	1.7	0.5	3.7	0.5	2.8	0.6
I experienced frustration while using the website	3.0	0.7	1.4	0.7	2.5	0.7
I experienced confusion while using the website	3.3	0.5	2	0.8	2.4	0.5
I found the website intuitive to use	2.6	0.7	3.3	0.5	3.1	0.3

Results from 3 groups

We converted the responses obtained in the form of "Strongly Disagree", "Disagree", "Agree" and "Strongly Agree" into the form of a scale ranging from 1-4. This is so that the data can be more easily interpreted, and is able to show the overall sentiment on the questions. Specifically, 1 now corresponds to "Strongly Disagree", 2 corresponds to "Disagree", 3 corresponds to "Agree" and 4 corresponds to "Strongly Agree".

- Participants using the prototype desktop were quite split on if they were satisfied with using the website or not. However, the sentiment weighed more heavily towards it being negative overall. In contrast, those using the redesigned sites on both methods overall agreed that they found the satisfying to use.
- Furthermore, participants on the prototype desktop site found the website overall not very usable with some participants still finding it having good usability but not many shared this opinion. On the redesigned sites all users responded positively to the question and said the site was good usability wise. A discrepancy does exist where on the desktop version it was much more highly agreed upon than on the mobile but it was still overall positive.
- A vast majority of participants for the prototype desktop site agreed saying that they experienced frustration while using the website and only a minority said they did not experience too much frustration. On the redesigned site, opinion is once again split for the redesigned site on desktop compared to mobile. A majority for the desktop site disagreed with the statement strongly and said they did not feel frustrated but for the mobile site most people said that they experienced some level of frustration throughout the usage of the site.
- When asked about confusion while using the website, a majority of participants on the prototype desktop site agreed they did feel confused at points during the test. When we compare this to the sentiment from the redesigned sites we see that it is more split; this is especially true for the desktop version where it is barely on a negative sentiment. Users were almost equally split on whether to agree or disagree with this statement. When we compare to the redesigned mobile site we get a slight favour to confusion but it is not by a large margin.
- Lastly, when asked on whether participants using the prototype desktop site found it intuitive

to use, the overall sentiment is that they found it intuitive but there a number of participants who strongly disagreed with this statement. When we compare this to the redesigned sites, all participants in both groups agreed that the site is intuitive to use and this sentiment is not split much and is the majority opinion.

In conclusion, the results of the survey show that participants clearly prefer using the redesigned websites when compared to the prototype website. This helps back up our earlier conclusion than the redesigned websites are easier to use. While they were some discrepancies in responses between the redesigned desktop and mobile sites they were not major, which supports our conclusion that the website is equally usable on both platforms.

3.2.2 Observations and Comments from Participants

While the participants were using the website some notes were taken on which features helped the participant and their overall comments throughout using the websites as well as their comments they offered at the end of the test. These comments and observations help further highlight the strengths and weaknesses of the features of the websites.

- We took count on how the users interacted with the search feature on the prototype website. Most commented on how they did not like that the search was not responsive and they had to confirm their search before they could see the item they were looking for.

  Conversely, on the redesigned sites most participants praised that the search bar was dynamic and appreciated that they did not need to type the whole item name before relevant results appeared.

  However, there were some bugs on mobile causing the search bar to behave unexpectedly, such as the results disappearing when pressing enter; this caused some frustration in participants, as they had to use the site with these bugs in mind.

- The cart button being separated from the top of the navigation and put as a floating button on the bottom right of the page received mixed responses from participants.

  On the desktop site, most participants did not mind the change and adapted to the location quite quickly. However, some participants still noted that they would have preferred it to have been more clearly labelled. Some participants however still liked the change and said they like this added simplicity of checkout and basket being combined into 1 click.

  When we compare it to the mobile redesigned site, we get similar responses where most users were able to adapt to the change quickly but a good chunk still preferred if it had remained in the navigation bar. One participant said that, at least for mobile, the intuition is that the checkout button should be in the hamburger menu.

- The categories feature we included in the redesigned sites also did not see much use. On the desktop site, most participants did not use the button and instead opted to scroll through the page instead to look for the relevant groups. When asked, most participants did not realise this was a feature.

  This is similar in sentiment for the mobile redesigned website, but one participant noted that they would have much preferred it if they had granular control of filters and sorting instead of just an implicit category feature.

- In the implementation of the add to cart feature on the redesigned sites, most participants found it intuitive to use (particularly on the desktop page).

  One participant did however experience frustration as they did not realise that add to cart was the same as setting the quantity of the item. This was a somewhat common comment participants had for the desktop version of the site, where they would have preferred if the wording of "Add to Cart" was changed to "Set Quantity" as it more directly reflects what is actually happening.

- For the redesigned website on mobile, most participants found it intuitive to use, but there was a common issue where if the participants double tapped the add to quantity button it would cause the page to zoom in. This was a source of frustration with some participants; some also wished that the item quantity could be changed to display 0 at the start.
- On the redesigned site and the prototype site, participants commented how they would have liked to have the capability to clear all items in the basket in a single press rather than do it manually per item.

  After discussion with the team, we also agreed that this is a feature that should have been implemented but was an oversight as we did not consider how many people would want this feature.

4 Limitations

4.1 Study Participants

Overall, while we were able to gather 9 participants for each category of our study (which was sufficient for our purposes), this meant that our results were susceptible to several sorts of **selection bias**, detailed below.

- Our results will have incurred **sampling bias**[6] to some extent, by virtue of most participants surveyed being university students from largely STEM backgrounds, meaning that they are likely to be more technologically adept than the average user and would be able to intuit the operations of user interfaces faster. Additionally, they all come from or are long-term residents of Western, educated, industrialised, rich and democratic societies[7], which may further shift the outcome of the responses just from considering a potential higher baseline of prerequisite knowledge.

  Given additional time and resources, we would ideally have used a larger sample size to be representative of a more demographically varied user base, including different age groups, ethnicities and backgrounds. This would better reflect the kinds of people most likely to be frequenting this website.
- **Self-selection bias**[8] was also a potential cause for concern, since the participants of the study were likely to already have been personally familiar with the assessors, and these relations may have caused skewed results.

  However, we mitigated this to some extent by ensuring that there was no external influence or power dynamic between assessor and participant that might affect the results, in some cases swapping planned participants across groups and assigned assessors to ensure this.

4.2 Study Format and Methodology

The structure of our study and how we conducted it inevitably opened up the possibility of **response bias** influencing the results alongside other impacting factors, which we will briefly discuss below.

- **Question order bias**[9] may have posed a potential issue, since our structure of questions was in a set format without factoring in randomisation. Positioning the collection of open comments after our Likert scale ratings may bias the quality of the comments themselves through priming; likewise, the positioning within the rating questions themselves may affect the ratings of those that follow it, especially since the first two have a positive implication.
- The fact that participants knew that they were being timed and would later be questioned regarding their use of the website may have unconsciously affected how they approached using the site; some seemed to try to complete their tasks as fast as they could, while others took their time in order to pick out aspects they could comment on, which may not reflect commonplace usage.

- In addition, **ecological validity**[10] is important to take into account; while ideally this study would have been performed in a controlled setting, due to time constraints and limitations on who we could call upon, we could not fully recreate a 'natural' setting in which the website could be used as if in full live deployment.

4.3 Coursework Restrictions

While the setting of the coursework helped to keep the team focussed with regards to our changes to the website, by restricting the work to only being a UI/UX update for the base prototype given, we were unable to have the freedom to undertake a complete overhaul of the website's contents (i.e. rebuilding the website from scratch).

One example was that some participants expected items to have more details such as a product description or branding on a separate page; writing these details for each item would have been well outside the scope for the project, even though this design would have been more accordant to potential customer expectations.

In addition, while exploring the website outside of the context of the controlled tasks we had our participants perform, some felt that the website felt lacking in terms of auxiliary content (such as in the checkout process or store details); there was little benefit to implementing something like this as it was in no way related to the actual user experience of the core website functionality as required by the coursework, so this was not taken as a priority. This included considerations for the performance and extensibility of the website.

While this lack of certain functionality and expected content does end up making the website feel 'unrealistic' despite being usable and consistent, these are unavoidable due to how the coursework is structured.

5 Conclusion

In order to obtain the most holistic results from our research we employed a multitude of methods to evaluate the different versions of the website, which also resulted in us obtaining a deeper understanding of the website. By performing A/B testing we were able to compare multiple versions of the website by measuring the total time taken to complete a set of tasks and we were also able to observe the areas of the task in which participants struggled with provided us a comprehensive view on the strengths and weaknesses of each version of the website. The usage of a questionnaire also allowed us to subjectively measure participants views and comments on the websites they used.

Looking at the results from the two t-tests we performed, we have found that on average, participants spend less time to complete the given tasks on the redesigned website when compared to the prototype website. We were also able to see that participants were able to complete the list of tasks just at least as fast on the redesigned mobile website as they would on the redesigned desktop website. The qualitative results we obtained also showed that the general consensus is that participants find the redesigned website on both platforms more usable compared to the prototype website.

Overall, based on the data we have obtained through our various methods of testing, we conclude that the redesigned website is more usable for users when compared to the prototype, and that the mobile version of the redesigned website provides at the very least equal usability to the desktop version.

References

[1] R. Procter, *Usability III: Evaluation With Users 1*, https://warwick.ac.uk/fac/sci/dcs/teaching/material/cs348/evaluation_with_users_1.pdf..

[2] M. J., *Usability metrics - a guide to quantify the usability of any system; 2019.* https://usabilitygeek.com/usability-metrics-a-guide-to-quantify-system-usability/..

[3] Y. K., *SPSS tutorials: Independent samples T test; 2013.* https://libguides.library.kent.edu/SPSS/IndependentTTest..

[4] K. W., *Analysis of variance (ANOVA) Explanation, Formula, and Applications Investopedia; 2024.* :https://www.investopedia.com/terms/a/anova.asp.

[5] S. J. S. University, *T distribution table*, https://www.sjsu.edu/faculty/gerstman/StatPrimer/t-table.pdf.

[6] *Survey Bias: Common Types of Bias and How to Avoid Them — qualtrics.com*, https://www.qualtrics.com/uk/experience-management/research/survey-bias/.

[7] J. Henrich, S. J. Heine, and A. Norenzayan, "The weirdest people in the world?" en, *Behavioral and Brain Sciences*, vol. 33, no. 2-3, pp. 61–83, Jun. 2010, ISSN: 1469-1825, 0140-525X. DOI: 10.1017/S0140525X0999152X. [Online]. Available: https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/abs/weirdest-people-in-the-world/BF84F7517D56AFF7B7EB58411A554C17.

[8] *An Overview of Survey Bias — surveycrest.com*, https://www.surveycrest.com/blog/survey-bias-sampling-types-prevention/.

[9] L. Chung, *The 7 types of sampling and response bias to avoid in customer surveys — delighted.com*, https://delighted.com/blog/avoid-7-types-sampling-response-survey-bias.

[10] S. Kieffer, "ECOVAL: Ecological Validity of Cues and Representative Design in User Experience Evaluations," *AIS Transactions on Human-Computer Interaction*, vol. 9, no. 2, pp. 149–172, Jun. 2017, ISSN: 1944-3900. [Online]. Available: https://aisel.aisnet.org/thci/vol9/iss2/4.

Appendices

A Script CS348 Script

Cem Yilmaz, Hawari Ansari, Damian Kwok, Viktor Saprykin

February 2024

Contents

1 Introduction

Welcome to the CS348 lab study script. If you are one of the conductors for the user-based website usability testing, please familiarize yourself with the following script, which outlines the procedures and guidelines to be followed during lab study sessions. Ask on Discord if anything is not clear.

2 Preparation

1. Ensure that all necessary materials for the study, including laptops, task lists, and consent forms(if wanted by the participant), are ready and accessible.
2. Familiarize yourself with the tasks participants will be performing and the overall study objectives.
3. Set up the designated areas for conducting the study, ensuring a quiet and distraction-free environment.
4. Verify that that the old website(given to the participant either on your own PC or sent as a file) and the URL for the redesigned version are working correctly.

3 Task Execution

Your main job would be to make the participants complete a series of tasks and record how long it took to complete all the tasks as well as give the participants a questionnaire (given below the task list) at the very end.

1. You should let the participant know that:
   - They'll be asked to carry out some simple tasks on a website
   - They'll be asked to complete a questionnaire
   - You'll be recording data about how they interact with the website Avoid using phrases like "our website" or "the old website", as this may influence the participant's perception of the website.
2. Obtain either verbal or written consent from the participant before continuing.
3. Provide participants with the list of tasks, given below, to be completed sequentially. Make sure the participant can see the shopping list on a separate screen (or piece of paper) to the screen they'll be using the website on. This is to make sure they won't spend time switching between the shopping list and the website.
4. Give the participant time to glance at the shopping list and make sure they understand it before continuing.
5. When the participant is ready to begin working through the task list, instruct the participant that as soon as the website loads they may begin. Load the website and start the timer.
6. Once the participant begins the tasks, refrain from interrupting or providing assistance unless absolutely necessary.
7. When participants indicate they have completed the final task, stop the timer and record the time taken to complete the tasks.
8. Encourage participants to vocalise their thoughts and experiences as they navigate the websites.
9. Monitor participants' interactions with the websites, noting any issues or difficulties they encounter.

4 Task List

1. Add 3 Artichokes to the basket
2. Add 3 fruits, 3 vegetables, and 3 other non-fruit/vegetable items to the basket
3. Use the search bar to add 4 salmons, an orange juice, and a radish to the basket
4. Add 5 pumpkins to the basket, go to checkout, and clear the basket
5. Add 2 breads to the basket, go to checkout, and add another bread
6. Add 2 plums and 2 chocolate bars to the basket, go to checkout, and remove the chocolate bars
7. Make sure the basket contains only 2 plums and 3 breads, then press "Pay" on the checkout page

5 Questionnaire

After participants have completed the tasks, provide them with either a series of verbal questions given below or a written questionnaire (with the same questions), depending on their choice.

1. Rate your satisfaction with the website on a scale from 1 to 5
2. The website good usability wise. Choose between: Strongly Disagree, Disagree, Agree, Strongly Agree
3. I experienced frustration while using the website. Choose between: Strongly Disagree, Disagree, Agree, Strongly Agree
4. I experienced confusion while using the website. Choose between: Strongly Disagree, Disagree, Agree, Strongly Agree
5. I found the website intuitive. Choose between: Strongly Disagree, Disagree, Agree, Strongly Agree

6 Debriefing

1. After participants have completed the questionnaire, if the participant is willing, engage in a debriefing session to discuss their experiences and gather additional feedback beyond the questionnaire.
2. Address any questions or concerns participants may have regarding the study or the websites.
3. Thank participants for their participation and valuable feedback.
4. Make sure the participant knows that their data can be withdrawn from the study at any time.

7 Data Collection

1. Compile the recorded data, including task completion times, questionnaire responses, debriefing comments and any observations made during the study.
2. Store all collected data securely in a shared location accessible to all study collaborators.

8 Conclusion

Thank you, Conductors, for your diligent efforts in conducting this lab study. Your attention to detail and professionalism contribute to the success of our research efforts. If you have any questions or require further assistance, please don't hesitate to reach out.

B Old Website Desktop Data

Time	Satifaction	Usability	Frustration	Confusion	Intuative	Additional comments
5:40	1	Strongly Disagree	Agree	Strongly Agree	Agree	Fix the basket not showing
5:22	3	Disagree	Agree	Agree	Strongly Disaagree	Search bar not responsive
5:05	1	Disagree	Agree	Strongly Agree	Agree	
4:56	3	Disagree	Agree	Agree	Agree	Can't close cookies
6:20	1	Strongly Disagree	Strongly Agree	Agree	Disagree	Clearing the basket manually was annoying
5:32	2	Disagree	Agree	Agree	Agree	
5:43	2	Strongly Disagree	Disagree	Agree	Agree	
6:01	1	Disagree	Strongly Agree	Strongly Agree	Disagree	No way to go back from checkout except back arrow
4:44	3	Disagree	Disagree	Agree	Agree	Intuative but things just not working

C New Website Desktop Data

Time	Q1	Q2	Q3	Q4	Q5	Comments
3:15	4	Strongly Agree	Disagree	Strongly Disagree	Agree	Cart not clearly labelled and would like to see a remove all from basket feature
2:50	4	Agree	Strongly Disagree	Disagree	Agree	Cart not clearly labelled and would like to see a remove all from basket feature
3:12	5	Strongly Agree	Disagree	Agree	Agree	Wants a mass delete from checkout button
3:32	4	Strongly Agree	Strongly Disagree	Agree	Strongly Agree	Search that shows vegetables first, confusing thinking there is no search match in different category, wants basket remove all
3:23	5	Strongly Agree	Strongly Disagree	Disagree	Strongly Agree	Checkout and basket is usually a seperate button but its simpler so not so bad
3:18	4	Strongly Agree	Strongly Disagree	Disagree	Agree	Categories not being seperate pages not standard, add quantity too ambiguous, should be changed to set to quantity
5:06	3	Agree	Agree	Agree	Agree	Add to basket is confusing and should be changed to set quantity
2:56	5	Strongly Agree	Strongly Disagree	Strongly Disagree	Strongly Agree	The cart would be better located at the top right of the nav bar
2:44	4	Agree	Strongly Disagree	Strongly Disagree	Agree	images for items could be better

D New Website Mobile Data

Total time (s)	Q1	Q2	Q3	Q4	Q5	Comments
216	3	2	3	3	3	Basket location intuition; not expected for mobile (needs at top), not visible enough. Screen elements scaled but not proportional. Search; results disappear on enter. Search wanted on basket screen, and always at top. Clear whole basket at once option.
295	3	2	3	2	3	Clicking on item shows up as 0; want to default to 1. Did not realise quantities could be changed at basket. Buttons, text small on mobile; issues with double tap. Zooming in to make buttons easier to click makes checkout button go away. Easy to remove items. Wanted sticky headers.
344	3	3	3	3	3	Wanted increment/decrement buttons at checkout. Weird zooming/scaling after zooming in and selecting elements.Responsiveness in general. Back buttons at checkout, titles on each page, ease of dynamic search. Look and feel of checkout is good but unclear. Hamburger menu scaling.
172	3	3	3	3	4	”Searching and pressing enter removed the search function. Buttons seemed a bit small as I had to zoom in to press some. Otherwise, felt very intuitive, and followed e-commerce shopping conventions. Thought the background was a bit too much contrast which was distracting. Checkout was expected to be in hamburger menu, but the floating button for the checkout was useful. Welcome Test User label as a button was unexpected. No logo on mobile. Still feels ””desktop site”””
208	3	3	3	2	3	Go to cart immediately option on item add confirmation. Cart in hamburger menu. Not mobile friendly in terms of scaling. (wanted 1 column), much zooming. Cart button disappears when zooming. Search bar too small. Was alright with quantity starting at 0. Remove all items option wanted (spam clicking leads to zoom). Much scrolling without use of search button, category jump buttons missing on mobile.Blank boxes with loading sign instead of display=none (i.e. placeholders when loading).
197	3	3	2	2	3	Tapping on items that are too small; zoom issue. No up and down to change quantity in checkout. Home is separate on the hamburger but would have been nice to have a logo. Clear filters and sorts wanted, rather than being implicit as categories. Checkout expected on top right or in menu. Search not as clear/intuitive. Rubberbanding as an issue.
311	3	4	1	2	3	Did not like item quantity starting at 0. Wanted clear all button on cart. Wanted item add confirmation to be centered at middle of screen, and wanted auto dismissal of this (like on the side, instead of having to click confirm each time). Expected search on the basket screen (white padding is misleading). Increment/decrement buttons on checkout. Wanted more clear navigation back from cart e.g. home or logo (avoid two clicks). Did not mind scrolling to find items.
152	3	3	3	3	3	Zoom issue on mobile, too heavy reliance on user zoom. Search makes things much faster, did not like scrolling. Jump buttons not on mobile, wanted there. Plus/minus buttons on cart items wanted. Zooming in and losing access to checkout button was a heavy gripe. Better indicators for loading images. Wanted product cards already seen to persist when scrolling. Item add confirmation causes screen to jump and blocks interaction. Wanted to possibly replace it with a popup/toast next to cart? Number of items on cart wanted instead of price.
260	3	3	2	2	3	Wish I didn’t have to zoom so much. Pressing add to basket as well as confirmation popup unnecessary, +- buttons at checkout, clear all items at checkout button wanted. Wanted item sorting and filters.