Analytic & Empirical Evaluation

Group 9: ShelfLife

Stakeholder Interaction

The concept behind our application was tackling Climate Change via the reduction of food waste at an individual level. We identified our users, which consisted of 18–24-year-olds, as the primary Stakeholder's of our application. Secondary stakeholders included farmers and food-distributors; however, they were inaccessible to us. Our stakeholder interaction was divided into the requirements phase and the design and development phase.

Requirements

We initially conducted interviews and sent out questionnaires, with the objective of obtaining further insight into our stakeholders' stance regarding our two areas of concern: Climate Change and combatting Climate Change via the reduction of food waste. To ensure a wide range of answers and avoid stakeholders answering with partiality, our initial questions were based solely on Climate Change. 62% of 21 surveyed, mentioned at least one food-related method to proactively tackle Climate Change, including avoiding beef or products with a high carbon footprint, reducing food waste etc. This reinforced our belief that tackling Climate Change via food waste was appropriate, while also showing that broadening the scope to include other food related concepts would be valuable.

Additionally, understanding how stakeholders tackled Climate Change meant we could integrate the relevant improved methods into our application. For instance, this led to making product carbon footprints traceable on an informatics page. This feature is particularly useful because the Stakeholders that expressed an interest in product carbon footprints also expressed uncertainty around the environmental impact of certain items.

It was then important to narrow the scope and delve into the specifics of food waste and other food-related concepts. Research demonstrated that "20% of food bought for households goes to waste" (Cuff, 2021) and "18- to 24-year-olds waste predominantly more than any other age group" (Nikolaus, Nickols-Richardson, Ellison, 2018), however 76% of 21 stakeholders surveyed "disagreed" or "strongly disagreed" to wasting a lot of food. So, if a user is oblivious to the amount food they waste, how are they expected to proactively try to reduce it? Even though stakeholder actions to combat Climate Change were predominantly food related, the need for additional incentives was necessary.

To formulate requirements that incentivised frequent use and appealed to users, whilst still adhering to our goals, we decided to try to further understand a user's meal planning habits and priorities when deciding what products to purchase. When planning meals, 81% of 11 interviewees mentioned they either, pick a recipe and buy the necessary ingredients, or base it on the ingredients they already have. This led to the emergence of two key features in the application: the inventory and recipe suggesting functionality. Having a virtual inventory with a user's ingredients and a recipe suggester (tool for suggesting recipes by optimizing the ingredients in a user's inventory), appeals to our stakeholders as it simplifies the meal planning process. Simultaneously, it satisfies our goal of reducing food waste, because users consume the ingredients they already have, rather than unnecessarily buying more.

When choosing food at the supermarket, stakeholders picked items based either on their needs, or on price and/or carbon footprint of the product. Therefore, to provide stakeholders with an additional incentive, we decided to add the ability to sort recipes in the recipe suggester by least additional cost and carbon footprint, however due to time constraints this was not ultimately implemented. Nonetheless, the recipe suggester indirectly satisfied these concerns, if less additional ingredients are bought, then that will tend to mean less additional cost and carbon footprint.

In addition, 55% of survey respondents ranked 'items unexpectedly expiring' as their top cause for food waste. This was also backed up by research – according to wrap.org.uk, just under half of avoidable food waste was “not used in time” (2020). This prompted us to include a notification system whereby users are notified of ingredients approaching expiry in their virtual inventory. This likewise appeals to our stakeholders as less food expiring means more money saved and a lower carbon footprint.

This iterative interview and survey process, allowed us to formalise our key application features:

- A virtual inventory with user’s items that kept track of expiry dates and alerted the user before a product expired.
- A recipe suggester that suggested recipes based on the items in user’s inventory.
- A personal informatics page that kept track of user’s product carbon footprint.

We carried out more interviews and questionnaires to collect ulterior feedback from our stakeholders on the newly defined features. 67% of the 12 surveyed stated that manually inputting expiry dates of food items was “inconvenient”. To simplify usability, we added automatically generated expiry dates calculable upon date of product purchase and expected shelf life.

Similarly, 58% of 12 survey responses demonstrated that manually inputting every item bought into the inventory was “inconvenient”. One stakeholder suggested implementing an automatically generated shopping-list from the items needed for a user’s desired recipes and other manually inserted products. This feature was not ultimately implemented due to time constraints, however, it led to the development of another important underlying feature. How was a user going to make a shopping list of generic products when shopping at a specific supermarket? This question led to the products in our application being supermarket specific, starting with Tesco products. This feature is useful as it means users can add the correct specific items bought to their inventory and allows for accurate data to be displayed regarding carbon footprint, as well as other relevant data we will make use of in the future.

Design and Development

Stakeholder engagement in the design phase was practically inexistant, primarily due to our naivety of what it entailed. We asked stakeholders to provide feedback on colour themes, completely dismissing important design choices made when designing the interface. This however was not the case in the application development phase.

Our development process involved a team of four programmers, all working together on a single given feature at a time, allowing us to prioritise essential features. More importantly, it allowed us to immediately start gathering stakeholder feedback on the application features in case major changes or improvements were necessary.

After a feature was developed, our stakeholders would carry out a set of tasks, so we could identify potential issues. We used the ‘Think Aloud Protocol’, having users report on what they were doing and thinking when carrying out the tasks. Additionally, we asked a series of open-ended questions to consolidate our understanding of potential issues. All the feedback collected was analysed using a rating schema based on frequency and impact of the issues, in order to prioritise fundamental issues.

In the inventory feature, it was evident that stakeholders struggled while adding items. Our initial design consisted of two separate tabs; one for adding items, one for displaying the items a user had. The connection between tabs was not obvious to stakeholders as they believed they were unrelated. Aside from stakeholders displaying obvious indecision when adding items, 22% of the 9 stakeholders were not able to complete the task.

In addition, 44% of the 9 stakeholders added items to the inventory twice, as it was not clear the items had been added. This was due to having to swipe from the adding tab to the inventory page, to see the items added. Therefore, we redesigned the inventory page, so that it was a single tab, displaying a user's items, with a 'plus' icon on the top right of the page for the adding of items. After this update, both unfamiliar and familiar users to the system immediately understood the purpose of the 'plus' icon. When the 'plus' icon is pressed, a popup tab appears, once the items have been added the popup closes, displaying a user's inventory as well as the items just added. This removes a user's uncertainty on whether he has indeed added the items to his inventory.

For the recipe suggester, the only notable feedback was obtained via the questions asked, where 44% of the 9 stakeholders indicated wanting additional recipe information, such as price and difficulty. The inclusion of this led to stakeholders requesting additional filtering and sorting mechanisms on suggested recipes. However, this was not implemented as it was not a priority in our rating schema, considering the limited time available.

The personal informatics page was simple in terms of navigation, as its main purpose was to effectively display data on carbon footprint. However, when asking stakeholders to provide insight on the meaning of data displayed, all except one demonstrated a lack of understanding on the impact of a user's average carbon footprint. Therefore, we included a Gauge meter, with segments colour coded from green to red, clearly highlighting the impact of a user's carbon footprint.

Once the application was developed, it was necessary to evaluate the system as a whole. The rest of the report explores the evaluation methods employed, as well as the evaluation of our application.

Task Analysis

Relevant Tasks

The 5 main tasks that we have identified within our system are the following:

- Adding Ingredients to Inventory
- Removing Ingredients from Inventory
- Providing Expiry Alerts
- Providing Recipe Suggestions
- View Environmental Data

Within this set of tasks, adding ingredients immediately stands out as the most important task, as the other four are dependent on ingredients being added to the inventory in the first place. This task is valuable to analyse, as we need to ensure that it is easy for users to keep an up-to-date list of their ingredients in the app, otherwise the expiry alerts, recipe suggestions and environmental data lose their effectiveness.

To support this decision, we asked a small group of stakeholders to use the app for a while, and then rank the tasks by which they thought they performed most often, and which they thought were the most effective at reducing food waste. Adding Ingredients had the highest mean rank of 1.6 out of all tasks for the first question, with the recipe suggester leading in the second question - also with a mean rank of 1.6.

Task Analysis Methods

To ensure that important tasks in our program are easy and intuitive for users, we decided to perform a hierarchical task analysis on 2 tasks. The reason we chose this technique for task analysis was that ease of use is crucial for our app especially - we need users to feel comfortable using it consistently to get them to reduce their food waste over time - and HTA diagrams are an effective way to evaluate the complexity of tasks and to find any areas where they can be simplified. Williams (2019) writes in a blog post, “it provides a quick illustration of how complicated (or simple) a task is”. In particular, we received a lot of suggestions from stakeholders on ways for users to add ingredients more easily to their inventory, and we believe an HTA should make clear where those methods could fit in.

Komninos (2020) writes about the benefits of task analysis, and how many improvements can be found “based on the simple hierarchy of tasks” while others “depend on a cognitive analysis approach”. He makes it clear that the analysis must include “a deep understanding for users, their environments, and their goals”. Therefore, we will also complete a cognitive walkthrough, and use those results to support our analysis. However, a hierarchical task analysis is especially useful in decomposing user goals that they are “notoriously bad at clearly articulating” into analysable subtasks.

Hierarchical Task Analysis

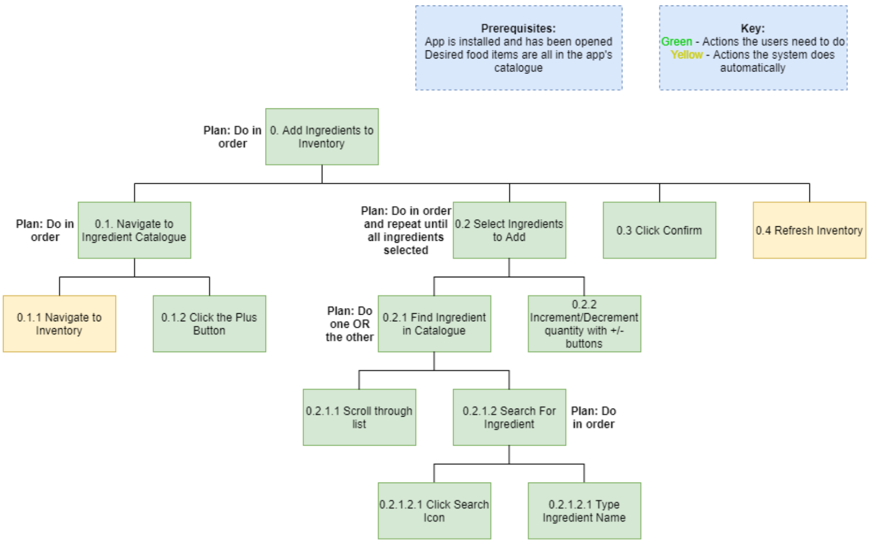

Figure 1.2.1: HTA Diagram for Adding Ingredients to Inventory

We have put a lot of thought into this task, and it already has many optimisations – for example, once found, adding ingredients takes only one button press for each. However, we can see that a lot of complexity comes with searching for specific ingredients – typing an ingredient name for each ingredient would be very time consuming and searching would be required if the catalogue of ingredients continues to grow. A simpler alternative here could involve a category system such that users can narrow down the catalogue to be small enough to scroll through and find their items easily. We could, for example, have categories for different suppliers, and diverse types of produce. This was never suggested by stakeholders and testers as our catalogue for the prototype was so small, and so this improvement is something we only could have learned from this task analysis process.

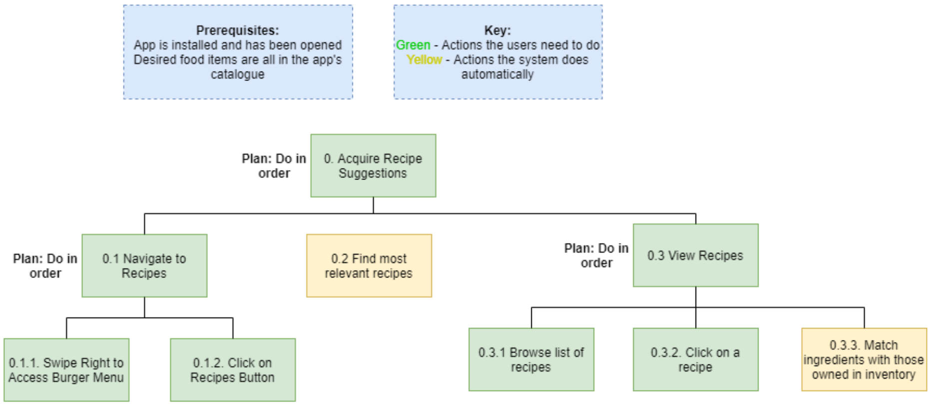

Figure 1.2.2: HTA Diagram for Providing Recipe Suggestions

This is the most technical feature of our system, so we are pleased with the number of clicks it takes to get to the recipes. One optimisation we could introduce in the future would again be the inclusion of categories to reduce the amount scrolling, such that the extensive list that might be hard to navigate can be grouped. The most complex part of this task is click on and off recipes so the user can view how many ingredients they currently own for each recipe, but the recipes are ordered so that the top suggestion is the one with the most ingredients owned. In the future, we would implement a recipe 'saving' system so that the user can have all their regular recipes available in one place, especially as the database of recipes and ingredients will grow rapidly.

Analytic Evaluation

In order the evaluate the usability of the system's interface we decided to use two evaluation methods, the first being a heuristic evaluation, and the second being a cognitive walkthrough.

The heuristic evaluation is an effective method to evaluate a system against a set of criteria, the heuristics; in this case, Nielsen's Ten Usability Heuristics for User Interface Design (Nielsen, 1994). Being simple and intuitive to perform, the heuristic evaluation is "easy to motivate people to do" (Nielsen & Molich, 1990). As it is performed by expert, system researchers identify usability issues with system's interface. However, as found in Nielsen and Molich's research (Heuristic Evaluation of User Interfaces, 1990), the percentage of usability issues found increases dramatically with the number of system researchers being assigned to review the system. With one researcher 20-50% of usability problems can be found whilst with ten researchers this increases to 71-97%. Whilst this indicates that the effectiveness of the review relies on the number of reviewers, due to difficulties because of the coronavirus pandemic, we decided to use three user researchers, which find, on average, 58% of usability problems. This is clearly an effective method for finding the majority of usability issues in a system which we can be used to improve and 'debug' ShelfLife.

In addition to the Heuristic Evaluation, we also decided to include a more formal method of evaluation, namely a cognitive walkthrough. Nielsen describes the cognitive walkthrough as "a more explicitly detailed procedure to simulate a user's problem-solving process" (Usability Inspection Methods, 1994). Compared to the heuristic evaluation, which assumes fundamental usability principles, cognitive walkthroughs provide a less biased view on the interface in the form of a stream of feedback from the user. This kind of feedback is valuable in our evaluation because it identifies specific issues and successes in our domain and application, as opposed to predefined heuristics. Using this information, we focus our attention to more critical problems rather than receiving an overview of the interface, which does not necessarily direct future work.

Heuristic Evaluation

To ascertain whether ShelfLife's interface met the appropriate standards of a usable interface, we asked three experts in user interface design to conduct a heuristic evaluation. To review the system, the system researchers were given time using ShelfLife on their mobile Android devices, as this is the only platform that ShelfLife has been developed for so far. The experts were asked to use the application until they felt comfortable that they had used and understood all the developed features of the application and experienced the interface in its entirety.

After spending time with the application, the system researchers were asked to discuss and judge the interface against established rules of interface design and assign a score for the application against each of these heuristics. The heuristics chosen to score the application against are Nielsen's Ten Usability Heuristics for User Interface Design (Nielsen, 1994). The full list of heuristics used can be seen in the appendix.

Results

The experts scored the system between 1 and 5 for each of the ten heuristics producing the results seen below.

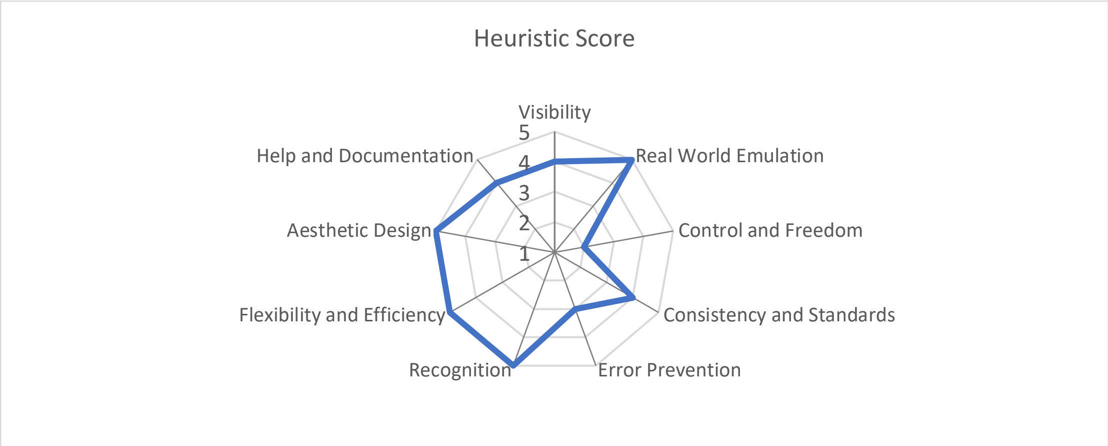

Figure 1.3.1 - Heuristic Evaluation Results

As well as scoring the heuristics quantitively, the experts were asked to comment on any aspects of the system they thought noteworthy. In reference to the high score (5) in the flexibility and efficiency of the system, the researchers mentioned that they found the system to be 'direct' and that 'each task is easy to complete'. This indicates that the interface of our system allows users to navigate and perform actions across the system with ease.

The experts found our system to be highly usable in general which will allow users to easily navigate and use the application. However, they found that the usability limited the users control and freedom, likely resulting in errors. The primary example they gave of this was the case where a user accidentally deletes an item from their inventory. While the experts though that 'using a bin [icon] for deletion is good', they thought an 'undo button' would be a worthwhile addition to recover from erroneous clicking.

Cognitive Walkthrough

As mentioned previously carrying out a cognitive walkthrough will provide further insight to tasks, and what critical problems arise at certain subtasks. This is achieved by asking ourselves 4 key questions, derived from Blackmon, Polson, et al.'s 2002 paper, "Cognitive walkthrough for the Web", whilst observing a user completing 2 main tasks.

The 4 main questions are:

- Will the user try and achieve the right outcome? Meaning: did the user understand the task they were asked to perform?
- Will the user notice that the correct action is available to them? Meaning: does the user realise they are able to perform the next action from the current screen?
- Will the user associate the correct action with the outcome they expect to achieve? Meaning: can the user recognise the next correct action from looking at the UI?
- If the correct action is performed; will the user see that progress is being made towards their intended outcome? Meaning: does the user recognise when they have performed the correct task?

We chose two tasks that encapsulate our program's features the most, but also coincide with the ones analysed in the Task Analysis, meaning we can get a well-rounded evaluation as mentioned in the Task Analysis methodology.
The steps to be taken by users to complete the tasks are:

- Adding onion to Inventory: navigate to the inventory, go to Ingredient menu, search "onion", Click on "+".
- Find remaining ingredients for Meatball Bolognese: navigate to Recipes menu, click on Meatball Bolognese.

Results

Before completing the walkthrough, we can conclude that the interviewees understood both tasks because no questions were asked about what they were expected to do, answering question one. The interviewees achieved the correct outcome for each subtask and didn't get lost when navigating the burger menu or carousel menu of the inventory page. This successfully answers the second and third questions of whether the users can recognise the next action they need to take and if they see the action intuitively executable from the UI.

While completing the first task the interviewees identified a bug, which was that the search bar doesn't disappear until 3 clicks, preventing the user from easily adding an ingredient, in this case an onion, to the inventory. This was critical for us to remedy before moving on to the empirical evaluation of the application as such an error would affect the results of these experiments.

To answer our fourth question, participants generally acted positively when they believed that they had completed both tasks, aside from participants who suffered from the search bar bug. Those who were affected tended to be surprised when the onion did not appear in their inventory after pressing the back icon to return as they had not pressed the confirm icon. Participants acted more confident in their positivity after completing the second task as they were more confident in their use of the system having previously correctly completed one task already. This demonstrates how easily learnable the interface's navigation is and how predictable the outcomes of a user's interactions are.

Empirical Evaluation

Usability of the System

Methodology

To have high confidence that our system is easy and quick to use, we gathered 3 sets of empirical data related to usability. The first set was the results of our main experiment. Here, we asked users who had never used the system before to complete two tasks using the app – the first was to 'add two onions to the inventory' and the second was ‘find which ingredients you need to make meatball Bolognese’. We timed the user completing this task and kept track of how many clicks they made. Then, these two numbers will be compared with the same set of users adding onions to their basket on the Tesco website, and finding which ingredients they are missing using the BBC Good Food website. Although not identical, we thought these tasks would be a good benchmark to compare to, as they are from popular websites which would have high usability standards themselves. We will compare the data using a two-tailed paired Student t-test, as we cannot rule out our system being worse than a competitor.

We chose these metrics as they are easy to measure accurately and give a good indication of the complexity of a task. Sauro (2011) says that in general, having less screens to move through will give a higher chance at task success, but criticises the method of only counting clicks, giving the extreme example of - “putting all functions on one or a few screens”. He goes on to explain that, although useful, counting clicks should not be a replacement for timing tasks. Therefore, we decided to measure both to make sure we have the right balance of intuitive screens and fast tasks.

Secondly, we will make note of the first click users make in these experiments. Research cited by Sauro (2011) states that on average, 87% of users will eventually succeed in a task when their first click is correct, which drops to 46% for users who clicked wrong. This justifies us performing a First Click Test, measuring the proportion of users that click the correct first button in each group of the experiment.

As for the third set of data, we had the users fill in a survey about the system’s usability. We took the format of the established System Usability Scale – according to Sauro (2011), “SUS has been shown to be more reliable and detect differences at smaller sample sizes than home-grown questionnaires”. This allows us to quickly evaluate qualitative feedback, even with a small sample size, which was all we could manage with social restrictions due to the pandemic and time restrictions. The survey consists of 10 questions (see Appendix 1), with the odd numbered ones asking for agreement with positive statements and the even numbered with negative statements. According to Sauro (2011), the questions encapsulate both usability and learnability.

Results

The p-value returned by the Student t-test for time taken to complete Task 1 versus the Tesco Website was 0.625, which does not show significant change either way. However, as for clicks, the p-value was less than 0.0001, showing a dramatic improvement over the competitor. As for Task 2, we saw a p-value of less than 0.0001 again for time taken, and 0.0008 for clicks, showing improvement in both categories here. (see Appendix 3). The first result shows that we have room to tailor the Add Ingredient task more to our app; as of now it is no faster than an arbitrary menu screen. On the other hand, the other results show us that the tasks are in general well optimised, and the lack of any extremely high outliers suggests that there is a low chance of a user getting stuck in any loops or giving up on a task.

100% of participants chose the correct first click for Task 1, being the plus button. We were confident that this result would be high as this specifically was one of the tests we performed with earlier prototypes of the system. In task 2, we had 6/7 participants tap the menu button, and one who tapped on one of their ingredients instead. These results show that the starting screen for the app is intuitive for new users and succeeds in leading them down the right path for the task they are looking to perform.

Using the recommended survey, our system achieved an average SUS score of 73.2 (see Appendix 2), which, being above 68, is above average according to Sauro (2011). This equates to a grade of B- and is an acceptable result (Sauro, 2018).

Effectiveness of the System

Methodology

We considered many methods for evaluating the effectiveness of our app. However, as most of the features of our app are suited to a long-term decrease in food waste, we thought a lot of our ideas would not show a significant enough change in the brief period we had to conduct the experiments. For example, the mass of a food waste bin can significantly fluctuate in a brief period due to different buying habits or events, despite decreasing slowly over a longer period. We decided that the way around this was to pick a feature and define a metric that would encapsulate its direct effect. The feature we thought would have the most significant effect on reducing food waste was the Recipe Suggester, and so the metric we chose was the mass of food used out of that which was bought in a week.

To conduct our experiment and to test the effectiveness of ShelfLife in reducing food waste, we contacted six potential users, asked them all to buy the same food from Tesco, the only supermarket ShelfLife is currently compatible with, and measure the mass of this bought food at the beginning and end of a week. Within these six potential stakeholders, half would be asked to use the app consistently to make food from the ingredients in their inventory, whilst the other half would not use the app. With the two masses from the start and end of the week from the two groups, each with three users, we will conduct a one-tailed Student t-test assuming normal distribution of the mass of food used to determine whether the app had a statistically significant effect on the mass of food not used in a week.

Using the data from our control groups, we will calculate the mean (μ) and standard deviation (σ) for the population for the distribution of the mass of food used. Using these, we construct the hypotheses:

- **H_0** – The data gathered from users of the app follows a standard distribution with mean μ.
- **H_1** – The mean is less than μ.

As well as conducting this statistical significance test, we will also ask those that used the app for a week to give feedback on their experiences using a survey. As the potential users who will have spent the most time with the app of all who have used it and used it for its purpose, these participants will have the best insight into the usability over a longer period.

Results

Participant	Group	Start Mass (Kg)	End Mass (Kg)	Mean	Standard Dev.	T-value
1	App Users	7.9870	1.0140	2.2330	1.0730	2.0033
2	App Users	7.9870	2.6510			
3	App Users	7.9870	3.0340			
4	Control	7.9870	3.2980	3.5935	0.4823	
5	Control	7.9870	4.1500			
6	Control	7.9870	3.3325			

Figure 1.4.1 - Effectiveness Experiment Results

A one tailed t-test at a confidence level of 80% yields a t-value of 1.886. Hence our t-value of 2.0033 allows us to conclude with 80% confidence that our null hypothesis – the mean mass of food left over between the two groups is the same – is likely not true. Therefore, we have evidence in favour of our application making a significant difference to the amount of food left over. Please note that we are not saying the metric of food left over directly influences food waste, however, we are assuming the two are related in the long term. Because of this, it is hard to conclude – over this period – that this result means our app tackles the problem of reducing food waste, which is the problem we aimed to tackle with this project. We assume that if people are using more of their food, then they are producing less food waste overall. Of course, providing evidence for this hypothesis would require much more data and a much longer time span to eliminate biases / anomalies such as longer expiry periods, people who eat more / less and seasonal occurrences etc. Furthermore, this result only had two degrees of freedom making it potentially unreliable. If we were to perform a better test, we would include more examples to have a better representation sample of the population. Due to feasibility and time constraints, we were only able to sample six households over a limited time, namely one week.

As well as this numerical data, we also found some other interesting trends within the experiment. For example, by the end of the test, a considerable amount of the extra mass used could be accounted for by meats and milk. While these are of course the higher mass foods typically used in meal preparation, they are also generally the quickest to expire. From this we can conclude that our app is causing users to use meat and milk because they are referring to it often for meal preparation or because they are the products that the expiry notifications will first warn them for (within the week of the experiment). Either way, our app is having a positive effect on the users use of these products in particular. Meat and milk also have a much larger impact environmentally than most other foods - in fact, dairy meat and eggs accounted for 83% of food waste emissions in the EU in 2018. (Ritchie and Roser, 2020), therefore our app reducing waste for especially these products makes a stronger contribution to our initial domain than other foods would.

Analysis & Future Work

In conclusion, the positive results from the effectiveness experiment, show that our app is successful in addressing the problem, albeit among a small sample. The usability experiment along with analysis of tasks and other analytical evaluation methods together provide strong evidence supporting the usability of our app, although those same methods have also highlighted areas in which our system could improve.

The results of our experiment showed that the app helps users reduce waste, particularly from highly perishable items. Therefore, we should ensure that our app stores accurate expiry dates for these types of items, however as of now they are generated automatically. This coincides with some stakeholder concerns that automatically generated expiry dates “could be dangerous”. To improve this, we will add a way to manually input expiry dates for instances of ingredients, even though expiry dates will still be generated automatically by default as not to greatly decrease usability.

The usability surveys were positive overall, and while a lot of people were happy with the Add Ingredient menu, others wanted a more convenient way to add food to the app. A few participants thought the process was “tedious”. This, along with the observations made from the task analysis about searching for specific items taking too long, motivates and justifies our idea of adding a receipt scanner to the app, allowing for automatic addition of an entire shopping trip's worth of items.

References

- Williams, T., 2019, *Hierarchical Task Analysis: Systematically Break Down Tasks into their Individual Steps [Online],* University of Bath Blogs. Available From: https://blogs.bath.ac.uk/ar-for-dementia/2019/01/18/hierarchical-task-analysis-how-to-systematically-break-down-tasks-into-their-individual-steps/ [Accessed 9 April 2021]
- Komninos, A. 2020, *How to improve your UX designs with Task Analysis [Online]*, Interaction Design Foundation. Available From: https://www.interaction-design.org/literature/article/task-analysis-a-ux-designer-s-best-friend [Accessed 9 April 2021]
- Sauro, J. 2018, *5 Ways to Interpret a SUS score, [Online]*, MeasuringU. Available From: https://measuringu.com/interpret-sus-score/ [Accessed 25 April 2021]
- Sauro, J. 2011, *Measuring Usability with the System Usability Scale (SUS) [Online]*, MeasuringU. Available From: https://measuringu.com/sus/ [Accessed 25 April 2021]
- Sauro, J. 2011, *Getting the First Click Right [Online]*, MeasuringU. Available From: https://measuringu.com/first-click/ [Accessed 26 April 2021]
- Ritchie, H., Roser, M. 2020, *Environment impacts of food production [Online]*, Our World in Data. Available From: https://ourworldindata.org/environmental-impacts-of-food#:~:text=When%20broken%20down%20by%20food,and%20dairy%20products%20they%20eat [Accessed 27 April 2021]
- Nielsen, J., 1994, *10 Usability Heuristics for User Interface Design [Online]*, Nielsen Norman Group. Available From: https://www.nngroup.com/articles/ten-usability-heuristics/ [Accessed 27 April 2021]
- Nielsen, Jakob & Molich, Rolf, 1990. Heuristic evaluation of user interfaces. *Proceedings of the SIGCHI Conference on human factors in computing systems*, pp.249–256. Available from https://bath-ac-primo.hosted.exlibrisgroup.com/permalink/f/1vdi69r/TN_cdi_acm_primary_97281
- Nielsen, Jakob, 1994. Usability inspection methods. Conference Companion on human factors in computing systems, pp.413–414. Available from https://dl-acm-org.ezproxy1.bath.ac.uk/doi/epdf/10.1145/259963.260531
- Cuff, M., 2021. *Almost 20% of food sold to homes, restaurants and shops ends up in the bin*. [online] inews.co.uk. Accessed from: https://inews.co.uk/news/environment/almost-20-per-cent-food-sold-homes-restaurants-shops-wasted-binned-898442 [Accessed 27 April 2021].
- Nikolaus, Cassandra J, Nickols-Richardson, Sharon M & Ellison, Brenna, 2018. Wasted food: A qualitative study of U.S. young adults' perceptions, beliefs and behaviors. Appetite, 130, pp.70–78. Accessed from: https://bath-ac-primo.hosted.exlibrisgroup.com/permalink/f/1vdi69r/TN_cdi_proquest_miscellaneous_2081549826

- WRAP, 2020. *Waste Prevention Activities – Food* [Online]. Available From: https://wrap.org.uk/resources/guide/waste-prevention-activities/food [Accessed 23/03/21].

Appendix

1 – Survey used for SUS score

The System Usability Scale

The SUS is a 10 item questionnaire with 5 response options.

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

The SUS uses the following response format:

Strongly Disagree 1	2	3	4	Strongly Agree 5
○	○	○	○	○

(Sauro, 2011)

2 – SUS test results

Participant	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	SUS Score
p1	3	1	3	1	5	1	3	1	5	4	77.5
p2	4	1	3	1	3	2	2	1	4	4	67.5
p3	4	2	4	1	4	1	3	1	4	3	77.5
p4	3	1	3	1	3	1	2	2	4	2	70.0
p5	5	1	4	1	4	1	3	1	4	3	82.5
p6	4	1	5	1	5	1	1	1	4	2	82.5
p7	3	3	4	2	3	2	3	2	3	5	55.0
											73.2

3 – Task time and click count data

Participant	Task 1 (Our System) Time	Task 1 (Benchmark) Time	Task 1 (Our System) Clicks	Task 1 (Benchmark) Clicks	Task 2 (Our System) Time	Task 2 (Benchmark) Time	Task 2 (Our System) Clicks	Task 2 (Benchmark) Clicks
1	35	39	9	13	25	45	4	10
2	39	36	10	15	18	51	6	9
3	39	42	7	12	20	36	5	12
4	35	36	7	10	21	41	4	8
5	40	43	6	10	36	49	4	13
6	40	39	9	14	19	40	4	12
7	41	38	8	10	22	48	6	9
Means:	38.42857143	39	8	12	23	44.28571429	4.714285714	10.42857143
SDs:								
Paired t-test P value	0.625		<0.0001		<0.0001		0.0008	

4 – Heuristics

- The visibility of system status,

- The match between system and real world,
- The user’s control and freedom in the system,
- The consistency and standards of the system,
- The error prevention protection in the system,
- The amount of recognition rather than recall needed to use the system,
- The flexibility and efficiency of using the system,
- The aesthetic and minimalist design,
- The help and documentation provided to recognize, diagnose, and recover from errors.

(Nielsen, 1994)