

TEST PLAN

STUDY OBJECTIVES

- 1 Identify weaknesses and strengths of the general UI navigation and structure.
- 2 Identify whether the users can easily find a recipe and buy ingredients.
- 3 Identify whether the users can easily find deals on the website.
- 4 Identify weaknesses and strengths of the multisearch tool.
- 5 Identify how satisfied users are with the website.
- 6 Identify if users can perform tasks in a more efficient way than on the competitors website.

RESEARCH QUESTIONS

- 1 What are the most common errors related to the general UI?
- 2 What are the completion rate of the tasks?
- 3 What is the time users need to complete the tasks?
- 4 How easily do users understand what is clickable?
- 5 How user navigate to complete tasks?
- 6 How users navigate between different sections and pages?
- 7 What are the mistakes?
- 8 How easily and successfully do users find the products or information they are looking for?
- 9 How well do users understand the symbols and icons? Which ones are problematic? Why?
- 10 Where in the site do users go to find Search? Why?
- 11 How easily can users return to the home page?

PARTICIPANTS

Participant ID	Participant Name	Researcher	Method	Test location	Device	Equipment / tool	Profile / Persona	Notes
P1	Arnold	Aga	Think Aloud	Home (in person)	Macbook	ScreenFlow	Persona - Andy	
P2	Tammy	Jill	Think Aloud	Home (in person)	Macbook	Microsoft Teams		
P3		Jill	Think Aloud	Home (in person)	Macbook	Microsoft Teams		
P4								
P5								
P6								
P7								
P8								
P9								
P10								
P11								
P12								
P13								
P14								

SCENARIO - The details that the moderator tells before starting the test, so that the participant has a better understanding of the context of use

Scenario 1	Description
	Imagine that you are a teller at Credit Union and you use a computer program to handle customer transactions. You are about to welcome your first customer today. The Credit Union system was updated recently and you are about to use it for the first time.

TASKS - The tasks assessed in this test

ID	Importance 1, 3, 5 or 8 (Fibonacci's scale)	Task	Expected Behavior
1	8	Please, describe what's on the screen.	N/A
2	8	Where would you look for help with finding an account?	User is able to Find Help Drawer
3	8	How would you find details of an account 194?	User is able to use Find Member Account Widget. Clicks into the input field and "194" is displayed.
4	8	You are about to serve your first customer. He used biometrics to enter the queue system. Please load your first customers account details to the screen.	User click "Start" to serve first customer. User clicks "View Details"
5	8	Describe what's on the screen. can you find members address?	User is able to see address in the Member Acc. details section
6	8	Your Customer would like to get a loan for €4000. Please Issue a loan to his account.	User click "Loan" button and click "Issue a loan" input field. Clicks "Confirm". Click "Confirm" again in the Warning Message window.
7	8	Your Customer changed his mind and wants a loan of €1500 instead. Please correct the amount.	User clicks "Reverse" just under success notification.
8	8	Where would you look for help how to correct that transaction?	User is able to find article about reversing transactions in the Help Drawer
9	8	What is the updated balance?	User is able to read updated balance.
10	8	Your Customer has two cheques. Please lodge these cheques into his current account.	User clicks "Lodge". Ignores Cash input fields and clicks into cheques input field. After fields are populated with data, user clicks Add Cheque button. Lastly, clicks "Confirm Lodgement" button, and "Confirm" warning popup.

SCOPE - Relevant system areas, modules and elements that are being tested (i.e.: screens, flows, UI elements, etc)

Name	Importance 1, 3, 5 or 8 (Fibonacci's scale)
Main Menu	5
Dashboard	8
Help Drawer	8
Advanced Acc Search	5
Account Details	8
Issue a Loan Form	8
Lodge Cheques Form	8
Reverse Transaction	8

ISSUE IMPACT LEVELS

Item	Value	Description
Suggestion	1	Just a comment, usually a suggestion
Minor	2	Participant stops to think, but proceeds
Major	3	Participant faces a significant delay or starts doing try-and-error
Blocker	5	Participant gets stuck or gives up, only proceeds with help

SUCCESS CRITERIA for Scoring Scenarios (Effectiveness)

Item	Value	Description
Success	1	<ul style="list-style-type: none"> • Completes the task with minimal effort • Reaches destination within 2 attempts • Does not receive hints from the facilitator • Does not encounter error messages • Does not mention frustration • Does not have suggestions for improvement

Partial Success	2	<ul style="list-style-type: none"> • Completes the task with moderate effort • Reaches destination within 3 attempts • Receives 1 hint from the facilitator • Encounters 1 or 2 error messages • Has to back up or reenter information • Has minor suggestions for improvement • Mentions minor frustration or expresses minor confusion
Failure	3	<ul style="list-style-type: none"> • Completes with considerable effort • Reaches destination in 4 or more attempts • Receives 2 or more hints from the facilitator • Encounters more than 2 error messages or the same error message more than once • Has to back up or re-enter information several times • Has major suggestions for improvement • Mentions frustration or confusion • Mentions they would have to call or speak with someone to complete the task • Concludes the task is complete, when it is not
Skip	4	<ul style="list-style-type: none"> • Does not complete the task • Gives up while trying to complete the task and concludes they cannot successfully complete it.
N/A		<ul style="list-style-type: none"> • Task skipped due to time • Not applicable - task not scored

Effectiveness	
	As a rule, the optimum respondent number for product effectiveness test is 11-15. This number of respondents is enough to reveal 90-95% of all major user errors in the product, statistical error of the result calculation being about 10%, so the overall product Effectiveness can be determined with sufficient degree of confidence: http://ui-designer.net/usability/effectiveness.htm
good	90-100%
normal	75-90%
bad	50-75%
awful	0-50%

SUS Scores - Descriptive Statistics of SUS Scores for Adjective Ratings			
SUS yields a single number representing a composite measure of the overall usability of the system being studied. Note that scores for individual items are not meaningful on their own. http://uxpajournal.org/wp-content/uploads/pdf/JUS_Bangor_May2009.pdf https://uixtrend.com/measuring-system-usability-scale-sus/			
Adjective	Grade Scale	SUS Score	
Excellent	A	> 80.3	
Good	B	68 - 80.3	
OK	C	68	67 - average score for a large web application
Poor	D	51 - 68	(Quantifying the User Experience: Practical Statistics)
Awful	F	< 51	

BENCHMARKS (time on task)										
Competitor	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7	Task 8	Task 9	Task 10

QUANTITATIVE DATA TO BE COLLECTED			
Type of data	Methods	Metrics	Notes
Task success rate	Conventional Usability Test		
Task completion rate	Conventional Usability Test	Effectiveness	Number of tasks completed successfully / Total number of tasks undertaken * 100%
Error rates	Conventional Usability Test	Effectiveness	Average number of errors per task
Time on task	Conventional Usability Test	Efficiency	Task Time = End Time – Start Time
Satisfaction questionnaire ratings	SUS (System Usability Scale) Questionnaire	Satisfaction	

QUALITATIVE DATA TO BE COLLECTED			
Type of data	Methods	Metrics	Notes
Observations about pathways participants took	Conventional Usability Test, Think Aloud		
Problems experienced	Conventional Usability Test, Think Aloud, Interview		
Comments/recommendations	Think Aloud, Interview		
Answers to open-ended questions	Interview		

USABILITY METRICS (The ISO/IEC 9126-4 approach to Usability Metrics)			
	https://usabilitygeek.com/usability-metrics-a-guide-to-quantify-system-usability/ http://ui-designer.net/usability/efficiency.htm		
Effectiveness	Example $E = ((3*1 + 1*0)/(4*1))*100\% = 75\%$		4 users work with a product according to 1 scenario. Three users complete the scenario successfully and one user fails.
Time-Based Efficiency	$Pt = (1/1 + 1/2 + 1/3 + 0/6) / (1*4) = 11/24$ (goals/sec)		*4 users work with a product according to 1 scenario.
Overall Relative Efficiency	$P = ((1*1 + 1*2 + 1*3 + 0*10)/(1 + 2 + 3 + 10))*100\% = 37.5\%$		Three users complete the scenario successfully and one user fails.
Expert Relative Efficiency	$Pe = 75\% * (4*1)/(1 + 2 + 3 + 10) = 18.75\%$		Scenario completion time with the first user is 1 sec, 2 sec with the second user, 3 sec with the third one, and 10 sec with the fourth (unsuccessful) one.
Task Level Satisfaction			
Test Level Satisfaction	SUS Calculator		

WE MIGHT NEED IT AT SOME POINT:

COMMENT TYPES - The way comments are categorized	
Item	
Positive	Praise, joy comment, etc
Negative	Complaint, reaction of annoyance etc
Neutral	Generic, indifferent comment (i.e.: Participant make a reinforcement of something trivial that happen during the test)

SYSTEM USABILITY SCALE

Use the table below to help you calculate the SUS score for each participant. A line has been filled out as an example.

Reference: <http://www.measuringusability.com/sus.php>

Participant	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 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.	SUS Score	Link to the Questionnaire Results
P1	5	2	4	1	5	2	5	1	4	1	90.00	
P2											50.00	
P3											50.00	
P4											50.00	
P5											50.00	
P6											50.00	
P7											50.00	
P8											50.00	
P9											50.00	
P10											50.00	
P11											50.00	
P12											50.00	
P13											50.00	
P14											50.00	
Value	Key										Result:	52.86
1	strongly disagree											Poor
2	disagree											
3	neutral											We are aiming for at least 68 (above average)
4	agree											
5	strongly agree											