



MSc in Computer Science

Team Project

User Evaluation Report

TrailSeek

Danish Jamil

Irfan Ali

Neel Salvi

Puja Verma

Shubham Jagtap

Vinicius Vieira

Andrea Curley

Damian Gordon

Paul Kelly

2020

1. Proposed Hypothesis

For the user evaluation, we decided to use two evaluation techniques, stakeholder unstructured interviews with a questionnaire and a usability test.

The focus of the questionnaire was to obtain qualitative and quantitative data. It was done to ten users who first tested the application and then were given a set of twelve open-ended questions for qualitative data. The questions were:

1. Do you have an existing app similar to TrailSeek?
2. How do you rate the app design?
3. Can you describe a situation in which the TrailSeek app is most useful?
4. Is our app helpful to achieve your goals of being active and social?
5. What goals is the app helping you achieve?
6. Would you like to add any function or feature to the app?
7. How would you rate the restriction and weather update features of the app?
8. Would you like any change in the position of any feature (on top or bottom)?
9. How would you rate our app overall?
10. What is your feedback on the design and colour combination of the app?
11. What is the reason for your score?
12. What can we do to improve?

And five questions with Likert scales to gather quantitative data. The questions were:

1. Do you understand the main goal of TrailSeek?
2. Do you understand the given important information related to Covid?
3. Do you understand how the App works?
4. Do you know how to Search Trail and Join, Create an event?
5. Are you familiar with the recommended trails and suggestions?

The focus of the usability test was to measure and quantify metrics in a controlled environment that could provide us with actionable insights. We have followed ISO 9241-11 recommendations since it provides a structured way to extract effectiveness, efficiency and satisfaction metrics from our tests.

According to Ohgi and Sauro (2011), the ISO 9241-11 standard defines usability as “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use”.

2. Experimental Method

2.1 Overview

The usability metrics for effectiveness are as follows.

- 1) Completion rate: a percentage on the number of completed tasks. According to Sauro (2010), the average task completion rate is 78%.
- 2) Number of errors: unintended actions, slips, mistakes or omissions a user can make while completing a task. According to Jeff Sauro, the average number of errors per task is 0.7.

The usability metrics for efficiency are as follows.

- 1) Time-based efficiency: the total time taken to complete a task.
- 2) Overall relative efficiency: the ratio of the total time taken by users who successfully completed the task in relation to the total time taken by all users.

The usability metrics for satisfaction are.

- 1) Task level satisfaction: a seven-step ordinal answer on how difficult/easy a task was to complete.

Following this technique, a set of six tasks was given to the user test group:

1. Search and find a specific trail
2. Create and view an event on a trail
3. Get weather information of a trail for the next two days
4. Join a trail event
5. Sign up with new credentials
6. Login in with existing credentials

2.2 Data Collection

Data collected during user evaluation included both quantitative and qualitative data and some background information of the user.

Qualitative data issues noted during the Stakeholder interview which were related to technical, usability and performance Issues were highlighted by users during the interview. Some interesting issues like Location permission were not handled, Nearby trails, easy trails buttons were looking inappropriate and many other issues related to visual representation were found by various users which wouldn't have been possible to find by us.

In addition, the responses obtained from questionnaires are also qualitative which is later turned into quantitative data for analysis purposes. A 0-5 Likert scale has been measured, Once it is

converted into the quantitative measure, the individual task success rate was calculated which directly relates to metrics used to evaluate the success rate of the system termed as Task Success rate.

Task success rate: It is used to measure if the users can complete the tasks successfully, this metric was used to evaluate the success rate of application. Quantitative data includes data collected using A questionnaire which derives insights about how much a user's expectations is achieved. Below are the metrics measured using the questionnaire.

Success/Fail - Number of Success or failure of task indirectly constitutes a good proxy for usability since every click or interaction should take the user close to their goal.

Time taken - Time spent on a page is another performance metric to be considered since it is an indicator of a false lead.

No. of Errors - errors can either slip; where the user accidentally types the wrong email address or picks the wrong dates when Inputting a birthdate, or they can be mistakes where the user clicks on an image that's not clickable or even double clicks a button or a link intentionally.

Task-level satisfaction - It is the completeness and accuracy with which users achieve specified goals. It is determined by looking at whether the user's goals were met successfully and whether all work is correct. This metric is a part of user journey success.

2.3 Selected Subjects

As part of the user evaluation, we had already conducted a study to narrow down our target audience which produced 4 personas for us. We tried our best to perform user evaluation on subjects who match up very closely with our personas.

As part of our stakeholder interviews, the User who matched our personas the closest were: -

1. User D - Persona - Abbey Keneddy
2. User E - Persona - Lilly Watson
3. User B - Persona - Mat Raul

The feedback from these users will be given more weightage moving forward as we implement the outcomes of the User Evaluation in the coming sprints.

For usability testing, a different set of users were used for data collection.

2.4 Data Analysis

The following table shows measuring criteria for the data collected through Usability Testing:

Task	No. of errors	Easiness of the task	Improvement Suggestion	Time in sec (Used to calculate app's efficiency)
Sign up	Success if 0 Fail if >0	Success ≥ 4 Fail < 4	Success if no improvements suggested Fail is improvements suggested	Time taken to complete the task. If a task is unsuccessful, time is recorded until the user gets an error.
Login	Success if 0 Fail if >0	Success ≥ 4 Fail < 4	Success if no improvements suggested Fail is improvements suggested	Time taken to complete the task. If a task is unsuccessful, time is recorded until the user gets an error.
Find specific trail	Success if 0 Fail if >0	Success ≥ 4 Fail < 4	Success if no improvements suggested Fail is improvements suggested	Time taken to complete the task. If a task is unsuccessful, time is recorded until the user gets an error.
Create and view event	Success if 0 Fail if >0	Success ≥ 4 Fail < 4	Success if no improvements suggested Fail is improvements suggested	Time taken to complete the task. If task is unsuccessful, time is recorded until the user gets an error.

Join event	Success if 0 Fail if >0	Success >= 4 Fail < 4	Success if no improvements suggested Fail is improvements suggested	Time taken to complete the task. If a task is unsuccessful, time is recorded until the user gets an error.
Get weather of next 2 days	Success if 0 Fail if >0	Success >= 4 Fail < 4	Success if no improvements suggested Fail is improvements suggested	Time taken to complete the task. If a task is unsuccessful, time is recorded until the user gets an error.

Table 1: Measuring criteria for usability testing

The above tasks were given to 8 users, and the team recorded no. of errors, ease with which user performed the task, any improvement suggestions given by the user while the user was performing the task and time taken to complete the task.

The responses collected from users were then assigned as Success (S) or Fail (F) for every task based on measuring criteria defined in the above table.

The analysis table is shown below:

Task	Metrics	User 1	User 2	User 3	User 4	User 5	User 6	User 7	User 8
Sign Up	No. of errors	1 (F)	1 (F)	1 (F)	1 (F)	1 (F)	1 (F)	1 (F)	1 (F)
	Easiness of the task	4 (S)	4 (S)	4 (S)	3 (F)	4 (S)	5 (S)	5 (S)	5 (S)

	Improvement Suggestion	Yes (F)	No (F)	Yes (F)	No (F)	Yes (F)	No (F)	Yes (F)	Yes (F)
	Time taken in sec	78	78	65	62	58	30	80	85
Log In	No. of errors	0 (S)	0 (S)	1 (F)	1 (F)	1 (F)	1 (F)	1 (F)	1 (F)
	Easiness of the task	5 (S)	5 (S)	5 (S)	5 (S)	5 (S)	3 (F)	5 (S)	5 (S)
	Improvement Suggestion	No (S)	No (S)	No (S)	No (S)	No (S)	Yes (F)	No (S)	No (S)
	Time taken in sec	17	20	20	23	17	130	10	10
Find specific trail	No. of errors	0 (S)	0 (S)	0 (S)	0 (S)	0 (S)	0 (S)	0 (S)	0 (S)
	Easiness of the task	5 (S)	5 (S)	5 (S)	5 (S)	5 (S)	5 (S)	5 (S)	5 (S)
	Improvement Suggestion	No (S)	No (S)	Yes (F)	No (S)	No (S)	No (S)	No (S)	No (S)
	Time taken in sec	50	50	46	41	32	54	20	30

[illegible]

	Improvement Suggestion	Yes (F)	Yes (F)	Yes (F)	Yes (F)	No (S)	No (S)	No (S)	No (S)
	Time taken in sec	10	10	10	9	15	10	7	5

Table 2: Analysis table based on measuring criteria

Based on the above analysis, the following conclusion was derived for each task.

Task	Completion Rate	Improvement Suggestion	Avg no. of errors
Sign up	0%	62.50%	1
Login	25%	12.50%	0.75
Find specific trail	100%	12.50%	0
Create and view event	100%	25%	0
Join event	100%	37.50%	0
Get weather of next 2 days	100%	50%	0

Table 3: Task performance measure

We see that Sign Up Page and Login Page have the least Completion Rate and most no. of errors, therefore these pages need to be worked on. We also see that 62.5% of users

suggested improvement on the Sign-Up page, 50% suggested improvement on Weather data and 37.5% users suggested change on Join event functionality. Therefore, these pages will be refined.

To measure the overall application's performance, we will measure 3 usability metrics(Justin Mifsud, 2015):

- Effectiveness: The accuracy with which users completed the task

$$Effectiveness = \frac{\text{Number of tasks completed successfully}}{\text{Total number of tasks undertaken}} \times 100\%$$

- Efficiency: The overall time taken by users to complete the task

$$Time Based Efficiency = \frac{\sum_{j=1}^R \sum_{i=1}^N \frac{n_{ij}}{t_{ij}}}{NR}$$

(Source)

Where:

N = The total number of tasks (goals)

R = The number of users

n_{ij} = The result of task i by user j; if the user successfully completes the task, then $N_{ij} = 1$, if not, then $N_{ij} = 0$

t_{ij} = The time spent by user j to complete task i. If the task is not successfully completed, then time is measured till the moment the user quits the task

- Satisfaction: Measure the overall ease of the use of the application

The following table shows the 3 metrics for the TrailSeek App.

Effectiveness(Task completion based)	70.83%
Efficiency(Time based)	0.04 tasks/sec
Overall Relative Efficiency	60.60%
Satisfaction(Task Level)	91.66%

Table 4: TrailSeek App Overall Performance

2.5 Practical Setup (200 words)

In practical terms how will you run your experiment? Will it be online or offline? What instructions will participants be given? What type of room/environment will the experiment be in? Will questions/surveys be displayed on-screen or on printed paper?

1. Keep the application steps instructions simple and make sure the application is working in perfect order at our end.
2. In an online experiment, test users will be displayed instructions and credentials required to run the app.
3. Install a necessary platform that is an “expo app” in users' devices and provide the related link and email and password to get access to TrailSeek. Any emerging issues will be dealt in a timely manner.
4. Steps to follow - After installing “expo app” from google play or apple App Store ->open the expo and login in to that with an email and password provided below -> open the link provided below -> Trailseek app will start.
Email: neel.salvi@gmail.com
Password: test12345
Link: exp://exp.host/@neels198/trailseek
5. Complete all tasks, the users will be given tasks they need to complete to test the app usability. The results of those tasks will be gathered and recorded for analysis.
6. Make a document of evaluation based on the feedback results gathered from the users through online and offline feedback.

If the user follows a step by step guide then reviews the project by the above principles, the user evaluation based on the questionnaire will be easy and user friendly in order to get the best possible output.

3. Conclusion

The outcome of the user evaluation highlighted 2 screens, login and signup, which need lots of improvement. The overall Usability testing score for the aforementioned screen was relatively lower than other screens and well as the stakeholders complained about the process which involved the 2 screens.

The User's closet to our personas had overall positive feedback about our app which suggests that we are moving in the correct direction. Our use of colour schema based on colour theory (Affairs, 2015) was appreciated by not only these 3 users but the rest of our sample users as well and the users could understand the main goal of the application clearly.

The Quantitative data gathered from the stakeholder interview also showed promising feedback. Although, the interview also highlighted some fallbacks of the application such as lack of testing on varying screen sizes, lack of screen space in some screen and lack of social sharing feature to name a few. Some of them are already in the pipeline for the app.

The sample set of users were lacking in the different type of users such as tourists and event organisers. A larger sample set with these types will further enhance user evaluation. A survey based on the Design Principles of Human Interface Guidelines by Apple (*Themes - IOS - Human Interface Guidelines - Apple Developer*, n.d.) will further allow us to enhance the user evaluation on a solid foundation.

References

Ohgi, J. and Sauro, J. (2011) [ISO standards and enterprise software](#): A case study using SUMI and SUS in an International Sale. In Marcus, A. (ed.), Design, User Experience, and Usability. Theory, Methods, Tools and Practice. Springer, Berlin.

Sauro, J. & Lewis J.R. (2010) “[Average Task Times in Usability Tests: What to Report?](#) .” in *Proceedings of the Conference in Human Factors in Computing Systems (CHI 2010)*

Affairs, A. S. for P. (2015, February 11). Color Basics. Department of Health and Human Services. [Color-basics.html](#)

Themes—iOS - Human Interface Guidelines—Apple Developer. (n.d.). Retrieved November 18, 2020, from

<https://developer.apple.com/design/human-interface-guidelines/ios/overview/themes/>

Ganglbauer, E., Schrammel, J., Deutsch, S., and Tscheligi, M. Applying Psychophysiological Methods for Measuring User Experience: Possibilities, Challenges and Feasibility. Workshop on User Experience Evaluation Methods in Product Development. August 25, 2009. Uppsala, Sweden

Alben, L. 1996, Quality of Experience: Defining the Criteria for Effective Interaction Design. Interactions, 3, 3, pp. 11--15.

Christou, G., Law, E., Green, W., Hornbæk, K. (2009). Challenges in Evaluating Usability and User Experience in Reality Based Interaction. Workshop in CHI'09.
<http://faculty.euc.ac.cy/gchristou/workshop/>

Ganglbauer, E., Schrammel, J., Deutsch, S., and Tscheligi, M. Applying Psychophysiological Methods for Measuring User Experience: Possibilities, Challenges and Feasibility. Workshop on User Experience Evaluation Methods in Product Development. August 25, 2009. Uppsala, Sweden

Vermeeren, Arnold & Law, Lai-Chong & Roto, Virpi & Obrist, Marianna & Hoonhout, Jettie & Väänänen, Kaisa. (2010). User experience evaluation methods: Current state and development needs. NordiCHI 2010: Extending Boundaries - Proceedings of the 6th Nordic Conference on Human-Computer Interaction. 521-530. 10.1145/1868914.1868973.

Hassenzahl, M., and Tractinsky, N., User Experience – a research agenda. In: Behavior & Information Technology, 25(2), (2006) pp. 91-97.

Usability Metrics—A Guide To Quantify The Usability Of Any System. (2015, June 22).

Retrieved November 18, 2020, from Usability Geek website:

<https://usabilitygeek.com/usability-metrics-a-guide-to-quantify-system-usability/>

Tullis, T., and Albert, B. Measuring the User Experience. Collecting, Analyzing, and Presenting Usability Metrics. (2008) Morgan Kaufman

Glossary

Lilly Watson



Goals

Lilly's long term goal is to advance he carrer and enhance her already impressive resume. Her shor term goal is to meet new people while alos improving her photography skills

Needs

Lilly want to find time away from her demanding job to pursue other personal interests.

Behavior

Regular commuter biker. No car. Interested in mountain biking but her friends don't do it

Interests

Graphic Design, Photography, Hiking, Camping, Socializing.

Fig 1: Persona - Lilly Watson

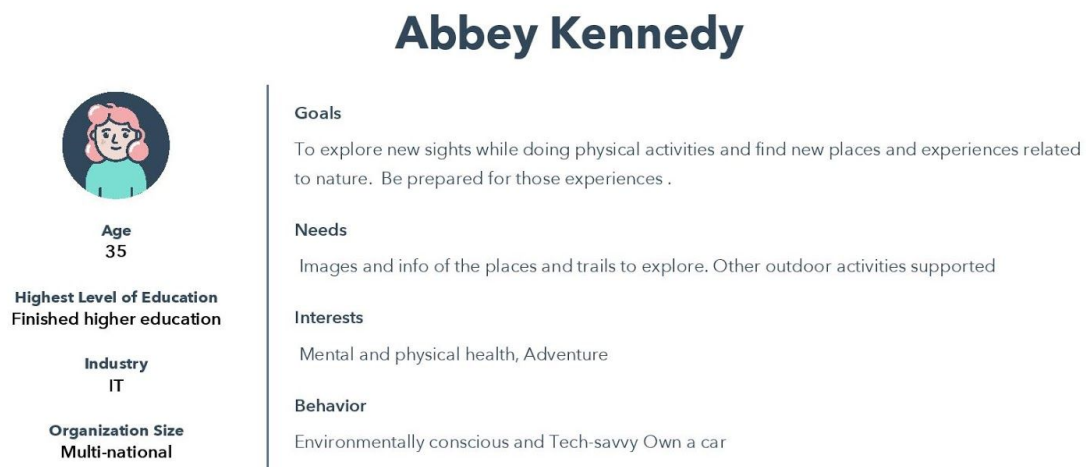


Fig 2: Persona - Abbey Kennedy



Fig 3: Persona - Mat Raul