Measuring UX – Two Examples

School of Computer Science | Software Analysis and Design
Marta Kristín Lárusdóttir, associate professor
7. October 2019

www.ru.is

Overview of this lecture

- Some Annoucements
- Types of Evaluation
- Measuring UX of Virtual Reality System
- Measuring UX of a Task Oriented Software

- Reading in Additional Material
  - ID book - Chapter 14.1 – 14.3
  - User Evaluations of Virtually Experiencing Mount Everest
  - Measuring the User Experience of a Task Oriented Software

www.ru.is

Results – Mid Term Evaluation

DCS-T-216-GHOH - Greining og hönnun hugbúnaðar

Are you generally pleased or displeased with the course?

5 Very pleased (9)	8%
4 Rather pleased (45)	39%
3 Moderate (37)	32%
2 Rather displeased (20)	17%
1 Very displeased (4)	3%
[Total (115)]	

Statistics	Value
Invited Count	332
Response Count	115
Response Ratio	35%
Mean	3.30

- Positive points
  - To have the lectures on-line
  - Good slides and well explained
  - Well organised
  - Good projects

Mid Term Evaluation – Possible Improvements

- Descriptions of the assignments
  - Could be more simple
  - Are time consuming
  - Some of the wording is not good
    - Please make notes in Piazza about text in the descriptions of the assignments that are unclear
- Have longer project solving classes
  - Actually you can choose in which class you attend
- TAs could be better prepared
  - I will explain better the assignments during our weekly meeting

In the News Paper – Thursday 3. October 2019

Success Rate - Árangursstig

- Success rate – author Jakob Nielsen
  - https://www.nngroup.com/articles/success-rate-the-simplest-usability-metric/

	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6
User 1	F	F	S	F	F	S
User 2	F	F	P	F	P	F
User 3	S	F	S	S	P	S
User 4	S	F	S	F	P	S

Note: S = success, F = failure, P = partial success

- In total, we observed 24 attempts to perform the tasks. Of those attempts, 9 were successful and 4 were partially successful. For this particular site, we gave each partial success half a point.
- In this example, the success rate was (9+(4*0.5))/24 = **46%**

Success Rate and Average Time Per. Task Problem Solving 4

- Overview

	Árangursstig	Tími
Opnunartími	100%	25.3
Upplýsingar um ákveðið vín	98%	34.2
Finna uppskrift	100%	61.7
Finna afhendingarmáta	95%	38.3
Finna vín undir 2000 kr.	80%	97.0
Útreikningur magns fyrir veislu	99%	59.7

- How would you react to these results?

Types of Evaluation

www.hr.is

Why, what, where, and when to evaluate

Iterative design and evaluation is a continuous process that examines:

Why: To check users' requirements and confirm that users can utilize the product and that they like it

What: A conceptual model, early and subsequent prototypes of a new system, more complete prototypes, and a prototype to compare with competitors' products

Where: In natural, in-the-wild, and laboratory settings

When: Throughout design; finished products can be evaluated to collect information to inform new products

www.id-book.com

10

www.ru.is

Why you need to evaluate

Bruce Tognazzini says:
"Iterative design, with its repeating cycle of design and testing, is the only validated methodology in existence that will consistently produce successful results. If you don't have user-testing as an integral part of your design process you are going to throw buckets of money down the drain."

See AskTog.com for topical discussions about design and evaluation

11
www.ru.is

Where Evaluation Takes Place

- Controlled settings that directly involve users (for example, usability and research labs)

- Natural settings involving users (for instance, online communities and products that are used in public places)

  - Often there is little or no control over what users do, especially in in-the-wild settings

- Any setting that doesn't directly involve users (for example, consultants and researchers critique the prototypes, and may predict and model how successful they will be when used by users

www.id-book.com   12
www.ru.is

6

Evaluation methods

Method	Controlled settings	Natural settings	Without users
Observing	X	X	
Asking users	X	X	
Asking experts		X	X
Testing	X		
Modeling			X

www.id-book.com 13

www.ru.is

Measuring UX – One Example
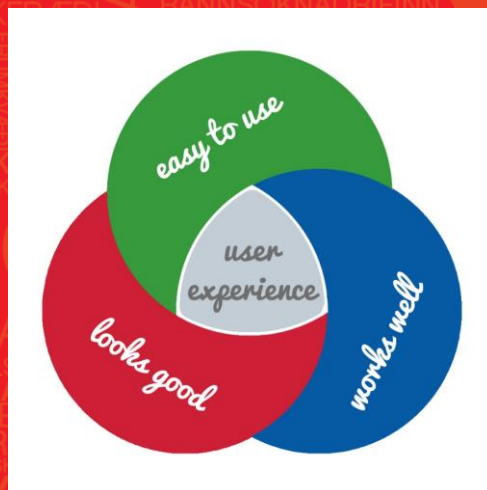
www.hr.is

Virtually Experiencing Mount Everest

- The company Sólfar wanted our advices of how to conduct user evaluations

www.ru.is

The User Evaluations

User evaluations:
- Introduction
- Short interview
- Pre-questionnaire
  - Simulator sickness questionnaire
- Experiencing the prototype
- Post-qestionniare
  - Attrakdiff 2.0
  - Simulator sickeness questionnaire
- Debriefing (through video)
- Thanking the user

www.ru.is

Participants

- User testing in one day
  - 5 participants – 4 males, 1 female
  - 25 – 54 years old
- Experiences
  - Had experienced VR before
  - Had some experience in games
  - One experienced in hiking

www.ru.is

Carefully Planned my Using RAMES

Roles	
R1. Users	Five users participated in the evaluation
R2. Evaluators	Conductor: Marta Larusdottir acted as a conductor
R3. Observers	Observers: David Thue and Kurt Van Meter acted as observers and assistants
R4. Recipients	Kurt Van Meter was the main recipient of the results
Activities	
A1. Purpose	To measure the user experience of the current prototype of *Everest VR*, to enable redesigning the system according to the results
A2. Plan	The user testing took place on Monday the 18th of April and Tuesday the 19th of April
A3. Evaluation procedure	1. Greet the participant 2. Short introduction to the procedure of the testing 3. Sign a consent form 4. Interview according to the background questions (Pre-questionnaire list) 5. Fill in the questionnaire about how the participant feels (Pre-questionnaire list) 6. Experience the VR prototype 7. Fill in the questionnaire about how the participant feels (Post questionnaire list) 8. Fill in the user experience questionnaire (AttrakDiff 2.0) 9. Discussion/debriefing about the video 10. Thank the participant
A4. Analysis of results	We used the Instant Data Analysis method described by Kellskov et al.[18]
A5. Making Decisions	Kurt was responsible for the decision making based on the results

www.ru.is

9

RAMES continued

Materials	
M1. Evaluation material	Pre-questionnaire kit including: a) introduction text for the participant, b) declaration of consent, c) pre-test questionnaire on the background, d) simulator sickness questionnaire. Post-questionnaire kit including: a) Post-test questionnaire on the overall feeling, b) simulator sickness questionnaire, c) the AttrakDiff 2.0 for measuring the user experience
M2. Support material	The VR prototype itself explained how to navigate between scenes in *Everest VR*. We also used a document containing an introduction to the procedure of the testing, and an introduction to the controls and the consent form.
M3. Data gathered	Background material, responses to questionnaires, usability problems, comments during debriefing sessions
M4. Results	Kurt presented the result to the team
M5. Decisions	Kurt kept track of which decisions were made

www.ru.is

RAMES continued

Environment	
E1. Evaluation environment	The evaluations were conducted at Reykjavik University, room M117
E2. Equipment. for data gathering	We used Camtasia to record what the user did during the VR experience
E3. Eq. to analyze results	Excel was used
System	
S1. Characteristics	VR game – Everest VR version 0.121
S2. Type	VR game
S3. Stage	Detailed prototype of the system
S4. Part	We evaluated the helicopter ride and a scene involving the Khumbu Icefall (part of the path up Mount Everest)
S5. Eq. for evaluation	Kurt provided all the equipment needed for the evaluation

www.ru.is

The Documents – Before the VR session

The Documents – Before the VR session

Virtually Experiencing Mount Everest

- The company Sólfar wanted our advices of how to conduct user evaluations

www.ru.is

Asked all users to explain what they experienced

www.ru.is

The Documents – After the testing

Instant Data Analysis

Problems Found According to Severity

Table 1: Number of User Problems in Each Severity Category

Severity Category	Number of problems	Average number of users
Showstopper	1 problem	5 users
Very severe problem	5 problems	3,6 users
Moderately severe problem	15 problems	2 users
Minor problem	9 problems	1 user
Total:	30 problems	

www.ru.is

What Was Done About It?

Table 2. Number of User Problems in Each Severity Category

Severity Category (number of problems)	Impact	Marked as useful
Showstopper (1)	Addressed	Very useful
Very severe problem (5)	4 addressed, 1 future design	1 very useful, 1 useful
Medium severe problem (15)	10 addressed, 2 no action 1 new tech needed, 1 not decided, 1 future design	4 useful
Minor problem (9)	4 addressed, 3 no action, 2 new tech needed	

www.ru.is

Some Quotes from the QA person

- The showstopper problem:
- *„This testing was the key in pointing out the importance of that"*

- One of the very severe problems:
- *„Huge impact on this from testing and we continue to reposition*
- *to find the best layout"*

- One of the medium severe problems:
- *„Being addressed, and was useful to have fresh eyes*
- *to underline the importance of this"*

www.ru.is

We Wrote a Paper on This

Hannes Högni Vilhjálmsson
Reykjavik University

David Thue
Reykjavik University

Marta Lárusdóttir
Reykjavik University

Paper title: User Evaluations of Virtually Experiencing Mount Everest,
- published at the HCSE conference, 2018.
- In Canvas as additional material

www.ru.is

Measuring UX – Another Example

www.hr.is

Measuring the UX of Task Oriented Software

- Paper presented at a COST 294 workshop in Reykjavik
- My Co-author was Jónheiður Ísleifsdóttir, master in computer science
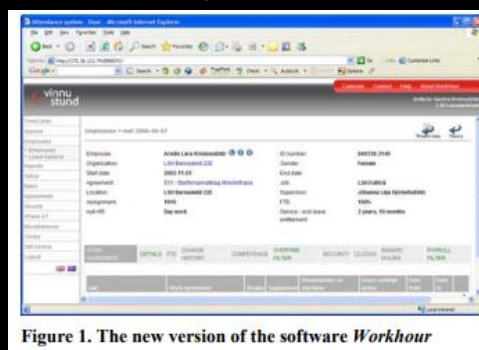- We did think-aloud testing on this system called Vinnustund

Figure 1. The new version of the software *Workhour*

www.ru.is

16

The Think-Aloud Tests

- We had 10 users – 5 ordinary users and 5 managers
  - 8 women and 2 men, working at the University Hospital or the Financial Management Authority
- We went to their work environment and did the user testing
- Wanted to focus on user experience before usage and during usage

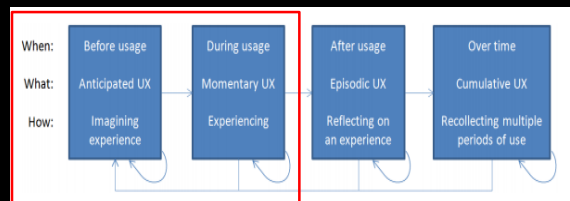| When:	Before usage	During usage	After usage	Over time
What:	Anticipated UX	Momentary UX	Episodic UX	Cumulative UX
How:	Imagining experience	Experiencing	Reflecting on an experience	Recollecting multiple periods of use

Figure 2. Time spans of user experience, the terms to describe the kind of user experience related to the spans, and the internal process taking place in the different time spans.

www.ru.is

Measuring the UX

- We used Attrakdiff 2 to measure the UX
  - Before the task solving session to measure their expectations
  - After the taks solving session to measure their actual user experience during the task solving session
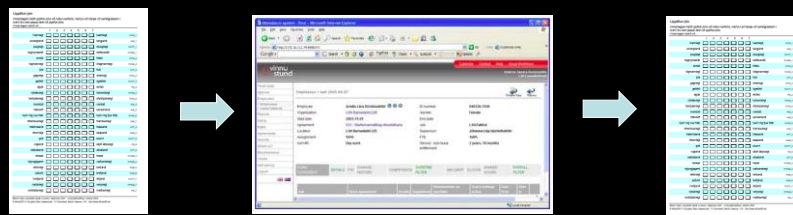
www.ru.is

Factors in Attrackdiff

- *Pragmatic Manipulation (PQ)*
  - Pragmatic attributes are the ones that makes us able to fulfill our goals and what we have until now talked about as usability
- *Hedonic Stimulation (HQS)*
  - The attributes encourage users to develop their skills and knowledge. These can be unused features of a software that the user does not yet use are not a part of the pragmatic experience but are rather perceived as Hedonic
- *Hedonic Identification (HQI)*
  - Hedonic identity attributes are the ones that make us identify with the product in a social context. These attributes are connected to the fact that all persons communicate their identity through things they use and own.
- *Attraction (ATT)*
  - Attraction is used to measure the global appeal of a product and to see how the other attribute affect this global judgment. We judge the product as a whole and use words like good, bad, beautiful and ugly to describe things.
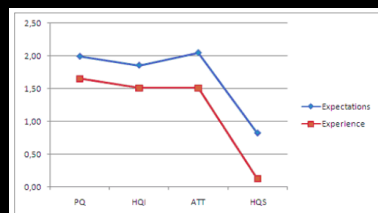
www.ru.is

Results

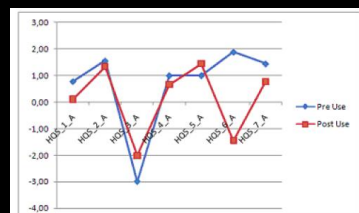Figure 2. Mean scores for each scale of AttrakDiff 2

Figure 3. Scores of HQS: Pre- and post-use

www.ru.is

Evaluation methods

Case	Mount Everest	Vinnustund
Why	Measuring UX in user testing. UX was measured right after usage	Measuring usability and UX in user testing. UX was measured before and right after usage
What	Nearly finished prototype – detailed design	Nearly finished prototype – detailed design
Where	In controlled settings	In the users work environment
When	2 months before launching	2 months before launching

Overview of this lecture

- Some Annoucements
- Types of Evaluation
- Measuring UX of Virtual Reality System
- Measuring UX of a Task Oriented Software

- Reading in Additional Material
  - ID book - Chapter 14.1 – 14.3
  - User Evaluations of Virtually Experiencing Mount Everest
  - Measuring the User Experience of a Task Oriented Software