



An Example Usability Testing

School of Computer Science | Software Analysis and Design

Marta Kristín Lárusdóttir, assistant professor

27. September, 2019

Overview of this lecture

- Measureable goals
- Comparing systems through user testing

- **Reading** (under Ítarefni in Canvas):

- Lárusdóttir, M. K., Ármannsdóttir, S.E. (2005) A Case Study of Software Replacement, Proceedings of the International Conference on Software Development, University of Iceland, Reykjavik, May 27 – June 1, 2005 pg. 129 – 140.

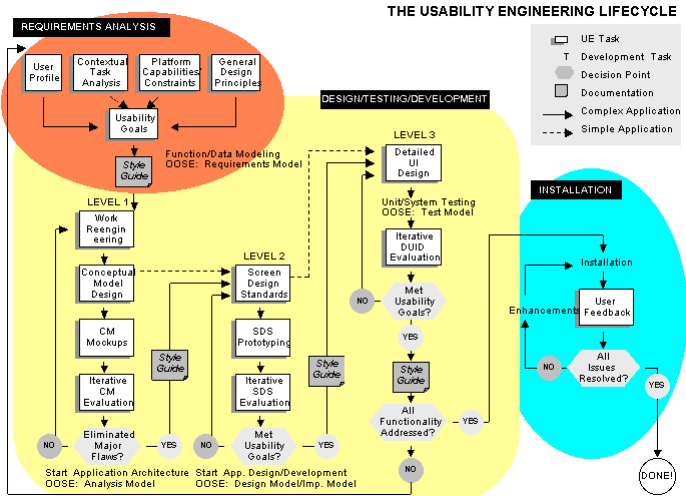


www.hr.is

Measurable goals



Why Measurable Goals?



Measureable Goals at Landspítali

STARFSÁÆTLUN LANDSPÍTALA 2019

LYKILÁHERSLUR

ÖRYGGISMENNING

- Öryggis- og umbátspjálftun starfsfólks
- Fyrirbyggja heilstu flokka atvika, sýkingar, byttur, lyfjastvik og þrýstingssár

ÁRANGURSVÍSAR

ÖRYGGISMENNING

Árangsvisar	Stofa 2018	Mánuði 2019
Hlutfall starfsfólks sem hefur fengið öryggisþjálfun	14%	40%
Fjöldi alvarlegra atvika	16	0
Spítalasýkingar	6,0%	5,8%

ÞJÓNUSTA

- Bæta aðgengi að þjónustu og stytta biðtíma
- Efta dag- og gönguvelðir
- Efting þjáknunar
- Auka þátttöku sjúklinga í eigin meðferð

ÞJÓNUSTA

Árangsvisar	Stofa 2018	Mánuði 2019
Hlutfall sjúklinga sem beðið hafa meira en 3 mánuði eftir meðferð	43%	40%
Meðalegultími, dagar	7,6	7,0
Ánægja sjúklinga með þjónustu spítalans (á kvarða 1-10)	8,3	8,5

MANNAUÐUR

- Bæta mönnum og vinnuskiptiag
- Bæta aðbúnað starfsfólks
- Efta leymsvinnu og framtýgja samskiptasáttmála

MANNAUÐUR

Árangsvisar	Stofa 2018	Mánuði 2019
Starfsánægja (á skala 1-5)	4,0	4,2
Ánægja starfsfólks með aðbúnað (á kvarða 1-5)	-	3,7
Ánægja starfsfólks með samskipti (á kvarða 1-5)	-	4,0
Hlutfall starfsmanna í 80% starfi eða meira	71%	74%

STÖÐUGAR UMBÆTUR

- Þróa og innleiða heildstæða ferla, frá upphafi til enda
- Styrkja alla stöðþjónustu í þágu sjúklinga og starfsfólks
- Styrkja umgöngu og skipulag vísindastarfs

STÖÐUGAR UMBÆTUR

Árangsvisar	Stofa 2018	Mánuði 2019
Ánægja starfsfólks með stöðþjónustu (á kvarða 1-5)	-	3,7
Fjöldi þetta greina í rýrindum tímartum	223	230

<https://stefna.landspitali.is/>

www.hr.is

Comparing Systems Through User Testing



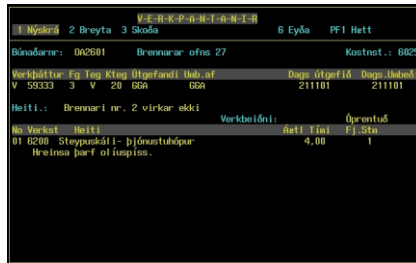
www.hr.is

The Motivation

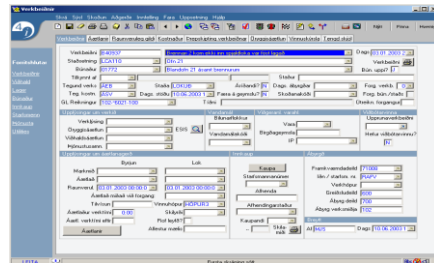
IT manager at a aluminum company asked:

“What impact does it have on my workers replacing their software system?”

“How can we measure that?”



The old system



The new system



www.hr.is

Measurements

Our approach was to measure the usability of:

1. The old system
2. The new system after 2 weeks usage
3. The new system after 6 months usage



We measured the

Effectiveness: “Can the user complete a defined task?”

Efficiency: “Can he/she complete it with reasonable resources?”

Satisfaction: “What are the benefits and drawbacks they mention?”



www.hr.is

Data Gathering Methods

Think aloud tests

- Users solved 6 predefined tasks at their work place
 - Made by a project manager
 - Adjusted to each user group
- One conductor
- One note taker
- The data:
 - Task completed or not
 - Usability problems
 - Users' comments
 - Data entered
 - How many systems used



The questionnaires

- Background
- After task, After test
 - The satisfaction



www.hr.is

Effectiveness - Completing tasks

Tafla 1: Samanburður lokiðna, ólokinna og verkefna með röngum gögnum

	CSD	Maximo (I)	Maximo (II)	Mismunur	
				Maximo II - CSD	Maximo II - Maximo I
Lokið	61%	38%	44%	-17%	6%
Rangt	22%	31%	36%	14%	5%
Ólokið	17%	31%	21%	4%	-10%

*Rauður litur = Maximo II í óhag.

Lokið = completed with correct data,
Rangt = completed with wrong data,
Ólokið = Not finished

CSD = old system,
Maximo I = new 2 weeks after installation
Maximo II = new 6 months after installation



www.hr.is

Effectiveness - Task Completeness

Table 1: The percentage of users completing successfully, completing with incorrect data and not completing tasks

User tasks	Completed with the correct data			Completed with incorrect data			Not completed		
	Old	New: short usage	New: long usage	Old	New: short usage	New: long usage	Old	New: short usage	New: long usage
Insert request -1	33%	0%	25%	67%	88%	75%	0%	12%	0%
Insert request -2	50%	0%	0%	50%	75%	100%	0%	25%	0%
Look up -1	50%	100%	75%	0%	0%	25%	50%	0%	0%
Look up -2	83%	75%	100%	0%	0%	0%	17%	25%	0%
Spare parts -1	50%	25%	38%	17%	17%	0%	33%	50%	63%
Spare parts -2	100%	25%	25%	0%	0%	13%	0%	75%	63%



Effectiveness Goals based on the Old System

- We build on the results from the user testing

User tasks	Completed
System used	Old
Insert request -1	33%
Insert request -2	50%
Look up -1	50%
Look up -2	83%
Spare parts -1	50%
Spare parts -2	100%

The format of Effectiveness goals:

More than < # > % of <user group> are able to complete < use case>

- The Goals

Usability goals - Effectiveness
More than 33% of all users are able to complete inserting request - failure
More than 50% of all users are able to complete inserting request - claims assessment
More than 50% of all users are able to complete looking up - check hours
More than 83% of all users are able to complete looking up - missing order
More than 50% of all users are able to complete finding spare parts - order a spare part
100% of all users are able to complete finding spare parts - quantity and placement



Are These SMART?

Usability goals - Effectiveness
More than 33% of all users are able to complete inserting request - failure
More than 50% of all users are able to complete inserting request - claims assessment
More than 50% of all users are able to complete looking up - check hours
More than 83% of all users are able to complete looking up - missing order
More than 50% of all users are able to complete finding spare parts - order a spare part
100% of all users are able to complete finding spare parts - quantity and placement



- Specific – Yes
- Measureable – Yes
- Attainable - ?
- Relevant – Yes (based on data)
- Time-bound – (Yes)

www.hr.is

Efficiency - Time on tasks

Tafla 5b: Samanburður á meðaltali tímamælinga

Verkefni	Maximo II - CSD			Maximo II – Maximo I		
	Maximo II	CSD	Mismunur	Maximo II	Maximo I	Mismunur
Stofna beiðni – Bilun	03:24	02:58	00:26	03:24	05:59	- 02:35
Stofna beiðni – Tjónsmál	05:38	03:39	01:59	05:38	05:32	00:06
Stofna beiðni – skrá vinnu	01:30	n/a	n/a	01:30	03:31	- 02:01
Uppfletting - Athuga vinnu	00:37	01:02	- 00:25	00:37	00:38	- 00:01
Uppfletting - Týnd pöntun	01:28	01:35	- 00:07	01:28	02:50	- 01:22
Varahlutir - Beiðni um varahlut	06:06	03:50	02:16	06:06	11:10	- 05:04
Varahlutir - magn og staðsetn	09:04	02:18	06:46	09:04	11:55	- 02:51

*Rauður litur = Maximo II í óhag.



www.hr.is

Efficiency – Measurable goals

Verkefni	Maximo II - CSD	
	Maximo II	CSD
Stofna beiðni – Bilun	03:24	02:58
Stofna beiðni – Tjónsmál	05:38	03:39
Stofna beiðni – skrá vinnu	01:30	n/a
Uppfletting - Athuga vinnu	00:37	01:02
Uppfletting - Týnd pöntun	01:28	01:35
Varahlutir - Beiðni um varahlut	06:06	03:50
Varahlutir - magn og staðsetn	09:04	02:18

*Rauður litur = Maximo II í óháð.

The format of Efficiency goals:

<user group> **are able to** < use case> **in** <limit> <resource> <on average/precisely>

Efficiency Goals
All users are able to insert request (failure) in less than 2:58 minutes on average
All users are able to insert request (claims assessment) in less than 3:39 minutes on average
All users are able to look up (check hours) in less than 1:02 minutes on average
All users are able to look up (missing order) in less than 1:35 minutes on average
All users are able to find spare parts (order a spare part) in less than 3:50 minutes on average
All users are able to find spare parts (quantity and placement) in less than 2:18 minutes on average



www.hr.is

Efficiency - Systems used

System used	Old	New: short usage	New: long usage
Insert request – 1	3,50	1,57	1,25
Insert request – 2	3,00	2,17	1,00
Look up – 1	1,00	1,00	1,00
Look up – 2	1,00	1,00	1,00
Spare parts - 1	3,75	3,00	1,37
Spare parts - 2	2,33	2,00	1,12

An efficiency goal based on these data:



All users are able to insert request (failure) switching less then 3,5 times between systems on average

www.hr.is

The Satisfaction

The old system

Benefits

I know how to use it

Drawbacks

Too many systems
Too much time

The new system

Benefits

One system
Easy to use

Drawbacks

Data is missing

**The users liked the new system better,
even though it was less effective (4 tasks) and less efficient**



www.hr.is

Conclusion in this study

What was the impact?

- More effective and efficient system
 - for solving 2 task, worse for 3, the same for 1
 - 2 task still bad efficiency after 6 months
- Users liked the new system better

The original question:

How can we measure the impact?

- With usability evaluation
 - Think aloud test and questionnaires
 - At the users work place



www.hr.is