# TDDD90 – Usability Testing

Användbarhetstestning

Shafiq Urréhman



Slides are adopted from Johan Blomkvist 2016-2017

#### Course info - Learning Outcomes

- Use methods and techniques for concept design and detailed design to define problems and alternative solutions for digital interactive products and services.
- Give an account of system objectives, and analyse design qualities and user experience for digital interactive products and services.
- Define purpose, content, and form for digital interactive products and services.
- Argue for one's interaction design ideas using multimedia, visualisations, or oral and written presentation.
- Summarise and analyse the meaning of concepts from interaction design and use them to analyse design work.



#### Where are we ...

#### Lectures

- Lecture 1, 14/9 at 10:15-12 <u>Course introduction</u> (Shafiq Uhrréhman)
- Lecture 2, 20/9 at 13:15-15 <u>Sketching details</u> (Stefan Holmlid)
- Lecture 3, 28/9 at 10:15-12 <u>Sketching concepts</u> (Mattias Arvola)
- Lecture 4, 11/10 at 13:15-15 <u>Ubiquitous computing</u> (Guest lecturer Dipak Surei)
- Lecture 5, 18/10 at 13:15-15 Creativity & Innovation in UX (Guest lecturer Dawid Ziobro)
- Lecture 6, 6/11 at 08:15-10 <u>Use Qualities</u> (Mattias Arvola)
- Lecture 7, 23/11 at 10:15-12 Usability testing (Shafiq Uhrréhman)



#### Where are we ...

#### • Seminars

Instructions about the seminars can be found <u>here (SWE)</u> or <u>here (ENG)</u>, and the seminar groups <u>here</u>.

- 30/11 at 13:15-17:00
  - Group 1-4 in R37, Group 5-8 in R44 Seminar 1: Paper prototype
- 13/12 at 13:15-17:00
  - Group 1-4 in R36, Group 5-8 in R44 Seminar 2: Pilot test



- Individual Assignment (project) Sketching & interactivity
- Read the full task first!
  - Concept Design Phase
  - Detail Design Phase
  - Prototyping
    - analysis
  - final Proposal



#### • Text from the task description ...

- Design Assignments Individual Continuation from assignment 2 (Grupparbete). A concept shall be further developed and detailed, prototyped and analyzed. The work is documented in the form of a description. **The description** is a chronological account of how the work has been done from start to finish and where it is stated: what ideas and alternatives have been explored and evaluated and what design decisions have been made and on what grounds.
- As a continuation of GroupWork/assignment 2, based on the ideas explored and the storyboard generated, a detailed solution to the assignment should be worked out and evaluated. Even ideas and concepts that were not included in the final proposal presented as a storyboard can be included and weighed in.



#### • key elements

- Begin sketching on a more conceptual level with inspiration from Group-Work assignment-2. Make sure not to do all sketch work on one go and at the same occasion, but rather spread over several instances.
- For the most important design decisions in detail design work, alternative solutions should be explored.



#### Detail design

For the most important design decisions in detail design work, alternative solutions should be explored.





Paper Prototype



Figur 23 - Pappersprototyp översikt



#### • Paper Prototype

The presentation/description 'format' plays a key role ...









#### Paper Prototype

- The presentation/description 'format' plays a key role ...
- There are mobile phone and other templates that you can print ...





#### **Prototype Seminars**

- Create a paper prototype of the final design proposal in accordance with Rettig (1994). The prototype should be displayed and discussed at seminar 1. Then make changes based on peers' feedback and create a more high resolution prototype. This version can be done by hand or digitally. The final prototype you will be tested/shown and discussed at seminar 2.
- Can be paper based or digital
- Text to support seminars and groups is available on the website





### Todays Lecture- Usability Testing

- A definition from ISO standard 9241
  - 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.
- Effectiveness (ändamålsenlighet):
  - accuracy and completeness in achieving goals
- Efficiency (effektivitet) :
  - resources expended...
- Satisfaction (tillfredsställelse):
  - comfort, acceptability (happiness, pleasure)



#### **Usability Testing**



#### Usability Testing vs Heuristic Evaluation





### Usability Testing vs Heuristic Evaluation





#### Questions

- Formative study
  - What are the most important usability issues?
  - What aspects of the product work well for the users? Is there anything that frustrates them?
  - What are the common mistakes users make?
- Summative study
  - Have the usability goals been achieved?
  - New product more effective than present?
  - Comparison with competing products?



## 'Roles' in testing

- Users Try to solve a given task by interacting with the prototype
- Computer knows the program logic and controls the interface, simulates the computer's response without comments

- Test Leader Manage the test session, give instructions to the user and ask for opinions and thoughts
- Observer notes silently



#### **Tasks Formulation**

- A reasonable number (lagom)
  - Representing, what the user is expected to use the system for
  - Max 1 hour, including surveys, interviews, etc.
- Give the tasks to the users one at a time and on separate paper
- On right level of detail
  - Not too many clues
  - Describe *what* the user must do, *not how*



#### Uppgift 1

Skapa en användare med ditt eget förnamn.

#### Uppgift 2

Ta fram ett matsedelsförslag utan att göra några egna inställningar.

#### Uppgift 3

Ta fram en ny omgång matsedelsförslag baserat på egna inställningar. Dina inställningar ska ta hänsyn till följande påhittade behov:

- Du vill ha matsedelsförslag för de kommande 4 dagarna.
- Förslaget skagälla för lunch (inte frukost eller middag).
- Du har älgkött i kylen som du gärna vill använda de kommande dagarna.
- Du tycker inte om vitlök och vill därför inte ha några recept som innehåller vitlök.
- Matsedelsförslaget ska ha god variation, och innehålla lättlagade recept.

#### Uppgift 4

Titta igenom de olika matsedelsförslagen. Sätt betyg på några av recepten baserat på hur goda de verkar vara. Ta fram en ny omgång matsedelsförslag och se vad dina betyg får för effekt.



#### Think About ...

• Ethical Issues

- Never defend the design to the user

- Practical preparation
  - Icebreaking
  - Training when needed
  - Always do a pilot study with 2-3 people



## Criteria for selecting participants

- Self-reported expertise
  - E.g. Beginner, Average, Expert, ...
- Frequency
  - Number of visits per month ...
- Amount of experience
  - Days, months, years
- Activities
  - Used special function



## Terminology 1

- Within-subjects
- Between-subjects
- Balance for (possible) learning between Tasks

Fp	Uppg 1	Uppg 2	Uppg3	Uppg 4
Fp1	U1	U2	U3	U4
Fp2	U3	U1	U4	U2
Fp3	U2	U4	U1	U3
Fp4	U4	U3	U2	U1



## Terminology 2

- Independent variable that which is being manipulated or controlled , for example
  - Characteristics of participants (age, gender, relevant experience)
  - Different designs-solutions or prototypes being tested
  - Tasks
- Dependent variable what you measure, for example
  - Task success
  - Time
  - System Usability Scale (SUS) score

- ...



#### Data Types

- Nominal (categorical)
  - For example, man, woman; Design A, Design B
- Ordinal
  - For example, ranking of 4 designs, from most beautiful to the least beautiful
- Interval
  - For example, 7-point scale of agreement: "This design is beautiful. Totally agree ... Do not agree at all "
- Ratio
  - For example, Time, Task Success%



#### Data Types (Cont..)

• Are these scales equal?

○ Poor ○ Fair ○ Good ○ Excellent
 Poor ○ ○ ○ ○ ○ ○ Excellent

- The upper is <u>ordinal</u>. Just calculate response rate.
- The lower can be considered as <u>interval</u>. You can calculate average.



## Usability Measurement

- Performance Measurement
  - Data success, time, error ...
- Problem-based measurements
  - Number of issues, type of problem ...
- Behavioral and psychological measurements
  Verbal behavior, facial expression ...
- Self-reported measurements
  - Expectations, System Usability Scale(SUS), ...



#### How / when should the measurement be used?

- Depends primarily on usability goals
- But, often applies:
  - In an early phase, task-related-success is most important
    - Task-related-success requires that serious problems are eliminated (*problem-based measurements*)
- Experience is important Does the user want to use the system?
  - Capture through behavioral measurements and selfreported measurements
- Time often has a threshold (but which?)
  - Should it be become Binary(below or above the threshold?)



#### Performance Measurement

- Task-related Success
  - Binary or different-levels
- Time
  - How long does it take for a task to complete
- Errors
  - Which errors or amount of errors are made per task
- Efficiency
  - number of clicks
- Learnability
  - How does performance change over time



#### Performance Measurement, Task-related Success $\rightarrow$ Binary

- Requires well-defined tasks with clear final terms
- "Find the price for the sofa 'Älmhult' in standard version."
  - OK?
- "Investigate different ways of retirement."

Participant	Task 1	Task 2	Task 3	Task 4	Task 5
P1	1	0	1	0	0
P2	1	0	1	0	1
P3	1	1	1	1	1
P4	1	1	1	1	1
P5	0	0	1	1	1
P6	1	0	0	1	1
P7	0	1	1	1	1
P8	0	0	1	1	0
P9	1	0	1	0	1
P10	1	1	1	1	1
P11	0	1	1	1	1
P12	1	0	1	1	1
Average	67%	42%	92%	75%	83%
Confidence					
Interval (95%)	28%	22%	29%	29%	29%





## To think about

- Reasonable threshold values?
  - Start from expert time , double it
- Handling outliers
  - Discount unreasonable times (long or short)
- Only successful-tasks or all tasks?
  - For failed tasks, if the user decided when to quit, use the time, otherwise don't
- Measuring time with think aloud?
  - Postpone extensive discussions to after the task has been completed
- Telling the user about the time measurement?
  - Ask the user to carry out all tasks as quickly and carefully as possible, without telling about the time measurement. If they ask, tell them that the start time and the end time is being noted



## Confidence intervals

• Suppose this is your data from task-related-success with 6 users

Försöks-	Uppgift 1
person	
Fp1	1
Fp2	1
Fp3	1
Fp4	0
Fp5	1
Fp6	1

- What does it mean?
- What does it mean for: how 'good' is your design?



### **Confidence** intervals

- Task-realted-success
  - discrete numbers, binary data
- Time measurement
  - continuous numbers, however, can be made to binary using thresholds



#### Number of "users"

- The purpose of the test
- Tolerance to error margin

Antal lyckade	Antal fp	Nedre 95% konfidensintervall	Övre 95% konfidensintervall
4	5	36%	98%
8	10	48%	95%
16	20	58%	95%
24	30	62%	91%
40	50	67%	89%
80	100	71%	86%



#### Number of Users



• Nielsen - 5 users identify 85% of user problems


# Number of Users

• "The short answer: Testing five users is not enough and magic numbers are strictly hocus-pocus /./ A mix of usability evaluation methods is most effective."

### Martin Schmettow -

http://www.utwente.nl/gw/cpe/en/Employees%20CPE/Schmettow/



## Efficiency

#### Combine task completion and time

Effektivitet —	Uppgiftsframgång
Ellektivitet =	Tid

Exempel:  $\frac{0.65}{1.5} = 0.4333...$ 

	Uppgiftsframgång	Tid (medel, i min)	Effektivitet (%)
Task I	65%	1.5	43
Task 2	67%	1.4	48
Task 3	40%	2.1	19
Task 4	74%	1.7	44
Task 5	85%	1.2	71
Task 6	90%	1.4	64
Task 7	49%	2.1	23
Task 8	33%	1.3	25



### Example

#### Efficiency (Task success per minute)





# Average effeciency (Tasks successfully completed per minute)





### Lostness

- N: Antal *olika* webbsidor som besöks under en uppgift
- S: Det *totala* antalet webbsidor som besöks under en uppgift
- R: Det minimala antalet webbsidor som måste besökas för att klara en uppgift

$$L = sqrt\left[\left(\frac{N}{S} - 1\right)^2 + \left(\frac{R}{N} - 1\right)^2\right]$$



http://bytes.schibsted.com/measuring-lostness/





# Usability Measurement

- Performance Measurement

   task success, time, error ...
- Problem-based measurements
  - Number of issues, type of problem ...
- Behavioral and psychological measurements
   Verbal behavior, facial expression ...
- Self-reported measurements
  - Expectations, System Usability Scale (SUS) ...



### Problem-Based Measures - What's a Problem?

- All as for anyone who is on wrong track
- Everything that creates confusion
- Everything that creates an error
- Not to see something that should be noted
- To assume that something is right when it is not
- To assume that a task is finished when it is not
- To perform the wrong function
- To misunderstand something/details
- Not understanding of the navigation



# Problem-based measurements--Details

- When does a problem start and end?
- Several observers?
- Granularity?



# Severity (Allvarlighetsgrad)

	Few 'fp' experiencing a problem	Many 'fp' experiencing a problem
Small impact on the user experience	Low severity	Moderate severity
Great impact on user experience	Moderate severity	High severity





#### Results - example



#### Antal unika användbarhetsproblem



#### Results - example





#### Results - example





#### Results - example





#### **Results - example**





#### Results - example





#### **Results - example**



Consequence in problem identification



## Thinks about – "Noise"

- Participants
- Information
- Method
- Artifact
- Settings
- Moderators



# Usability Measurement

- Performance Measurement
  - task success, time, error ...
- Problem-based measurements
  - Number of issues, type of problem ...
- Behavioral and psychological measurements
   Verbal behavior, facial expression ...
- Self-reported measurements
  - Expectations, System Usability Scale (SUS) ...



### Behavior and psychological measurements

- Verbal behavior
- Facial expressions
- Eye-tracking
- Pupil Feedback
- Heart Rate
- ...



### Behavior and psychological measurements

• Eye-tracking



https://www.filfak.liu.se/presentation/namnder/ikt/iktstudion/kontakt?l=en



# Usability Measurement

- Performance Measurement
  - task success, time, error ...
- Problem-based measurements
  - Number of issues, type of problem ...
- Behavioral and psychological measurements
   Verbal behavior, facial expression ...
- Self-reported measurements
  - Expectations, System Usability Scale (SUS) ...



## Self-reported measurements

- Expectation Dimensions
- <u>System Usability Scale (SUS) scale</u>
- <u>The computer system usability questionnaire (CSUQ) scale</u>
- <u>Questionnaire for User Interaction Satisfaction(QUIS)-scale</u>
- ...



- Ask participants ('fp') *about expected severity before performing the task*
- Ask how easy/hard it really was, immediately after each task
- Use the 7-point scale in both cases

- Very easy to very difficult



ask 1. Find out how many people report to Jeremy George's manager. (Include Jeremy In this number.) How Difficult or Easy do you expect this task to be? Very Difficult 〇 〇 〇 〇 〇 Very Easy	
Task 2. You need to return a notebook to someone but only know that he works on the 4th floor of 245 Summer St (building code V,floor 4), and that his manager's name is Tom. Find this man and report his corplD.How Difficult or Easy do you expect this task to be?Very Difficult OOOOVery DifficultOOOVery DifficultOOOVery Easy	
Task 3. You remember talking with someone named John and want to contact him, but don't remember his last name. You only know that his last name starts with S. You also remember that he works at 500 Salem St. in Smithfield (Building code OS) and work in FIIS. Find this man and report his corpID.         How Difficult or Easy do you expect this task to be?         Very Difficult Or Corpore Corp Very Easy	is
ask 4. You know a woman who works in FISC whose first name is Edyta and you need to call her. What are the last four digits of her phone number? How Difficult or Easy do you expect this task to be? Very Difficult 〇 〇 〇 〇 〇 Very Easy	

🚳 Tasks 1 - Microso	dt Internet Explorer						
Task 1 of 7:	7: Find out how many people report to Jeremy George's manager. (Include Jeremy in this number.)						
	Enter answer:	Overall, this task was:					
		Very Difficult C O O O O Very Easy	NEXT Task >>				
-							

- For each task, calculate two mean values
  - Average for the expected value
  - Means of experience value
- Visualize data as a two-axis scatter plot
  - expected value
  - experience value
  - The four quadrants are interesting





1=Difficult; 7=Easy



### Self-reported measurements→ SUS- System usability scale

- <u>SUS</u> Developed at Digital equipment corporation(<u>DEC</u>)
- Consists of 10 questions
- Can replace "website" against "system" against "product" ...



## Self-reported measurements→ SUS- System Usability Scale

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



## Self-reported measurements→ SUS- System Usability Scale

1. Jag tror att jag skulle vilja använda denna produkt ofta.

1	2	3	4	5				
Instämmer inte alls				Instämmer helt				
2. Jag tyckte att den	na produkt v	ar onödigt komplicer	ad.	J. Jay lyckle all del	аннь юг шу	นพละ แม่งปาวลหงลา	s i produkteri.	
1	2	3	4	1	2	3	4	5
Instämmer inte alls				Instämmer inte alls				Instämmer helt
3. Jag tyckte att den	na produkt v	ar lätt att använda.		7. Jag kan tänka mig	g att de flesta	skulle lära sig		
1	2	3	4	att använda denn	a produkt my	/cket snabbt.		
Instämmer inte alls				1	2	3	4	5
				Instämmer inte alls				Instämmer helt
<ol> <li>Jag tror att jag kol för att kunna anvå</li> </ol>	nmer att bei inda denna j	növa hjälp av en tekr produkt.	nisk person	3. Jag tyckte att den	na produkt v	ar mycket besvär	lig att använda	l.
1	2	3	4	1	2	3	4	5
Instämmer inte alls				Instämmer inte alls				Instämmer helt
5. Jag tycker att de o	olika funktion	erna i denna produk	t är väl san	). Jag kände mig vä	ldiat trvaa nä	ir iag använde de	nna produkt.	
1	2	3	4	4	2	2	4	5
Instämmer inte alls				Instämmer inte alls	2	3	4	o Instämmer helt
				10. Jag behövde lära	a mig myckei	t innan jag kunde	komma igång	med denna produki
				1	2	3	4	5
				Instämmer inte alls				Instämmer helt
			l					



### Self-reported measurements→ SUS- System usability scale

- tain a total value. The values for individual
- SUS results in a total value. The values for individual questions say nothing.
- Calculation:
  - Each question's value varies between 0 and 4
  - Questions 1, 3, 5, 7, and 9 contribute with the position minus 1
  - Questions 2, 4, 6, 8 and 10 contribute 5 minus the position
  - Multiply the sum for all 10 questions by 2.5 for the total value.
- SUS values vary between 0 and 100



### Self-reported measurements $\rightarrow$ SUS-System usability scale





SUS Score= 22 \* 2.5 = 55

### Self-reported measurements→ SUS- System usability scale

- Why SUS ?
- **Task Success :** Overall, users of the NASA site got **58%** of their tasks correct while users of the Wikipedia site got **71%** of their tasks correct,





https://www.measuringux.com/Apollo/index.htm

### Self-reported measurements→ SUS- System usability scale



% of "Correct" Conclusions

Figure 2. From "A Comparison of Questionnaires for Assessing Website Usability," by T. S. Tullis and J. N. Stetson, 2004, Proceedings of UPA 2004 Conference. Reprinted with permission.



ullet

### Eye-tracking video










## Eye-tracking video

### Eye Tracking

- Learn how people interact with the UI by watching where they look
- Two types used:
  - Heat map
  - Gaze tracking
- Help to understand linguistic similarities and differences
  - Right-to-left reading languages
  - Vertical text languages



# **Further Read**

- <u>https://www.interaction-design.org/literature/topics/usability</u>
- <u>http://edutechwiki.unige.ch/en/Usability\_and\_user\_experience\_surv\_eys</u>
- <u>https://www.nngroup.com/articles/usability-101-introduction-to-usability/</u>
- <u>https://www.researchgate.net/publication/235850976</u> Efficient Mea <u>surement of the User Experience of Interactive Products How t</u> <u>o use the User Experience Questionnaire UEQ Example Spanish</u> <u>Language Version</u>



#### Slut

#### www.liu.se

