

Evaluating and adjusting - 18.03.2016

Time	Subject
8.30	Agenda and a recap from last time
8.40	Evaluation and testing (Qualitative research)
9.05	A/B split test – clickstream and more
9.40	Questionnaires and surveys
10.00	Break
10.30	Expert analysis with Heuristics
11.20	Sit down and user test
12.00	Homework and happy holydays.

Todays Schedule



Journey map workshops

- Visually illustrates customers' processes, needs and perceptions via a thorough scenario.
- Map to understand & diagnose experience issues.
- Use maps to reframe and reimagine experiences.
- Redesign experiences to influence attitudes.
- Scenarios are powerful tools for idea creation, evaluation and understanding.
- Attitudes drive behaviours that deliver results. An experience derived from behaviour can change an attitude, thus influencing results.

Attitudes drive Behaviors

Behaviors deliver Results

Experiences Influence
Behavior



Last time



This time:

Topic: Evaluating and adjusting. Quantitative research. A/B split test, Clickstream, Surveys and usability testing with Heuristics.

Read / Do:

[Unger] Russ Unger et.al: *Find out how they think* p. 279 - 310 [Cooper] *On validation and testing* p. 139 – 143 [Snyder] *On paper prototypes. Alternative to the slow prot.* p. 138 – 154

Check out:

[Schneider] Bens Schnidermans 8 heuristic principles for usability [Nielsen] Jakob Nielsens 10 heuristic principles for usability

This time





Quantitative



Other common contrasts

Quantitative	Qualitative		
Numbers	Words		
Point of view of researcher	Points of view of participants		
Researcher distant	Researcher close		
Theory testing	Theory emergent		
Static	Process		
Structured	Unstructured		
Generalization	Contextual understanding		
Hard, reliable data	Rich, deep data		
Macro	Micro		
Behaviour	Meaning		
Artificial settings	Natural settings		

Qualitative and quantitative



Quantitative answers:

- What users do?
- How they do stuff?
- When and how often?



- Where they move and where they are active?
- How old are users, where do they live, what education and job?
- Do they prefer A or B and where do they look?
- Are users happy or unhappy with your product or service?

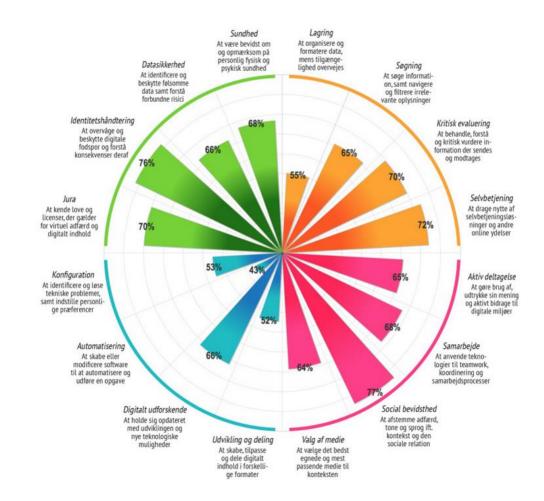
What can quantitative research tell you?



"Data mining" to find the pattern behind.

Trying to find the **why** of a certain behaviour.

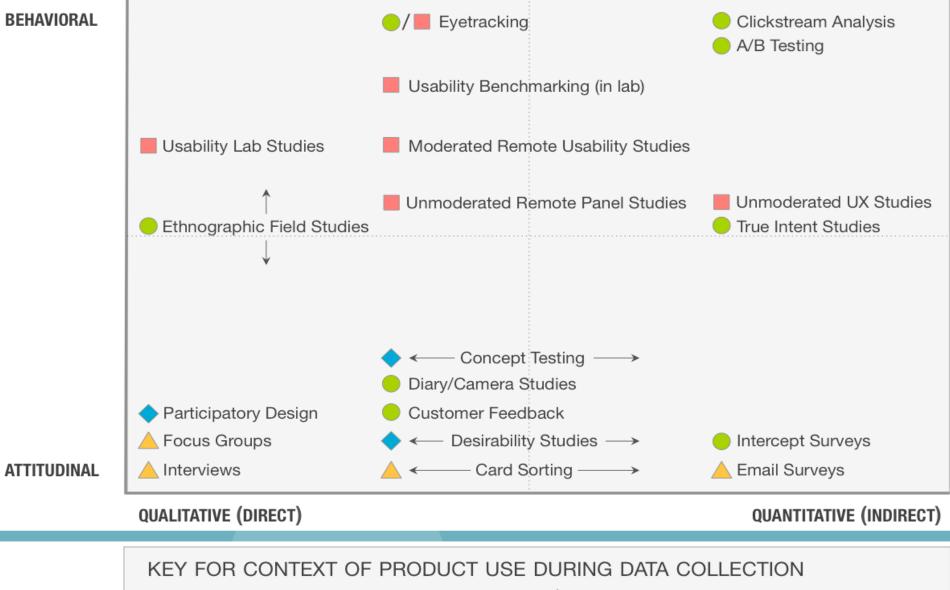
Useful when the **why** is to abstract or the behaviour is social rather than personal.



Digital Literacy



A LANDSCAPE OF USER RESEARCH METHODS



© 2014 Christian Rohrer Natural use of product
Scripted (often lab-based) use of product

De-contextualized / not using productCombination / hybrid



What is the difference?

When and how did you work Quantitatively?

What did you gain by doing so? Was it useful?

What questions did the research answer?

Two types of research

Evaluation and user testing – are you worth your price?

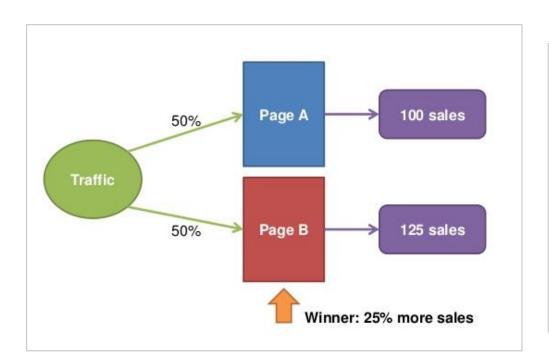
We will look at:

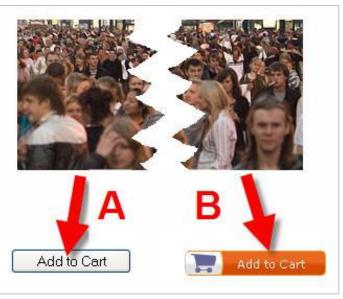
A/B Split testing
Statistics and heat maps
Heuristics
Sit down usability test
Questionnaires

"Usability testing is, at its core, a means to evaluate, not create"- Cooper et.al p. 140

Why?





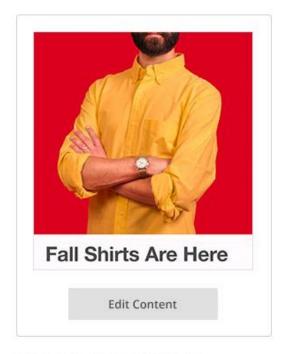


A/B split testing



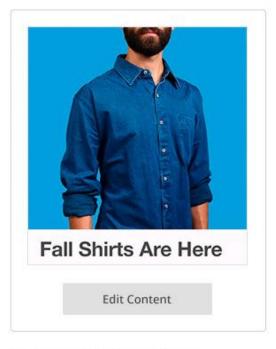
Content setup

Test different messages or templates within your campaign.



1 Column - Banded Template Edited on Aug 07, 2015 09:48 am

"Testing the order in which the content appears: main alpha image (red), headline,"



1 Column - Banded Template Edited on Aug 07, 2015 09:37 am

"Testing the order in which the content appears: main alpha image (blue), headline,"



1 Column - Banded Template Edited on Aug 07, 2015 09:53 am

"Testing the order in which the content appears: main alpha image (yellow), headline,"

Example

Early Online Access List

Export 🗸

15,000 Recipients

List: Early Online Access

Tested on: 40% of 37,500 subscribers

Variates tested: Content

Subject: Fall Shirts Are Here

Subscribers per combination: 5,000

Delivered: August 6, 2015, 3:00pm

Winning metric: Open rate

View Emails · Download · Print · Share

Overall open rate	42.7%	Overall click rate	21.3%
List average	45.3%	List average	20.4%
Industry average (Ecommerce)	27.1%	Industry average (Ecommerce)	13.7%

Winning combination



|LNAME|, Fall Shirts Are Here

Sent Aug 06, 2015, 3:00 pm from Shelby (freddie@freddiesjokes.com)

"Testing the order in which the content appears: main alpha image (blue)"

5,000 Sends

45.7% Open rate 22.8% Click rate

View Report

Example



Almost anything on your page that affects visitor behavior can be A/B tested

- Headlines
- Sub headlines
- Paragraph text priority
- Content near the fold
- Call to action text
- Call to action button
- Links (color and size)
- Images
- Testimonials
- Awards and badges
- Media mentions

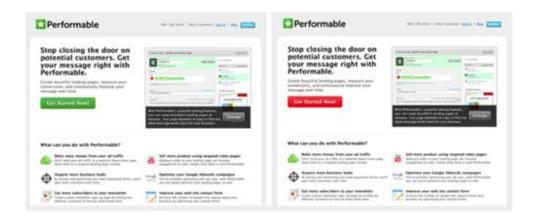


A/B split





Conversion up 34%



Conversion up 21%

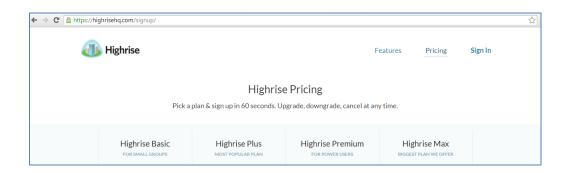
A/B case studies





A/B case studies







From

"Start a Highrise account"

To

"30-day Free Trial on all Accounts"

30% increased conversion

A/B case studies on text





Traffic logging found out:

Visitors abandon the purchase process at the time of checkout.

A/B split found out:

A single page checkout works much better at completing sales.

A/B case studies



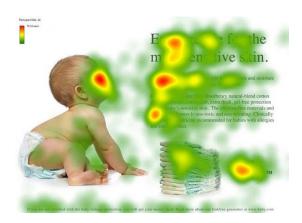
We recognize and react to faces on Web pages faster than anything else.

Faces looking right at people will have the greatest emotional impact on a Web page. The eyes are the most empathy triggering part of the face.

If a face on a Web page looks at another spot or product on the page, people will also tend to look at that product.



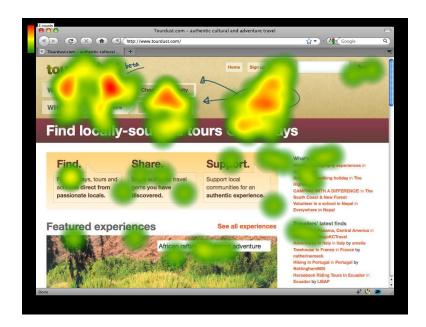




A/B case studies



Can be quite good – but ask yourselves how you would use the results? You discover people look here – what do you do?





Heat map or eye tracking



Make sure:

- 1. Not to end the test to early! You are better of with a 1000 participants than 100 remember statistic validity.
- 2. Run the test for a longer period of time. You have to rule out changes in conversion rate due following weekdays and paydays.
- 3. To plan several test to cover the month and year.
- 4. Have enough traffic. If you don't have enough traffic don't waste time testing. Make the changes and change back if it dos not work.
- 5. You test a hypothesis! Don't just test random stuff and different layouts. You have a hypothesis when you have a proposed statement made on the basis of limited evidence that can be proved or disproved.

How?

Make sure:

- 7. You do not spend time on simple color testing alone. Change the visual hierarchy colors don't matter on there own.
- 8. To be happy if you discover even small gains! A few percent can make a huge difference.
- 9. You know validity: Are you sure your tools is working correctly? and are you sure you understand the test results? Is anything happening in the world that might affect you test(political, holyday etc.)? Did you only test a selection of you users(newsletters)? Does your code work on all platforms?
- 10. You don't run to many tests, at the same time with overlapping traffic.

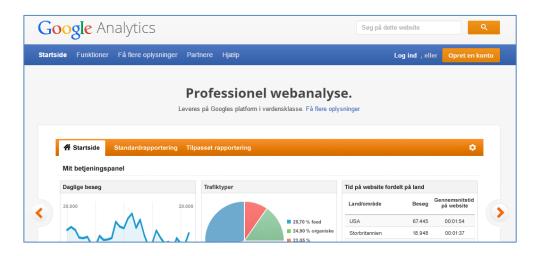
How?

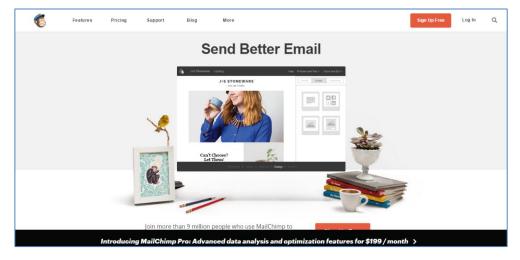
Tool for web and apps:

- Google analytics
- Optimizely

Tool for newsletters:

- Mailchimp





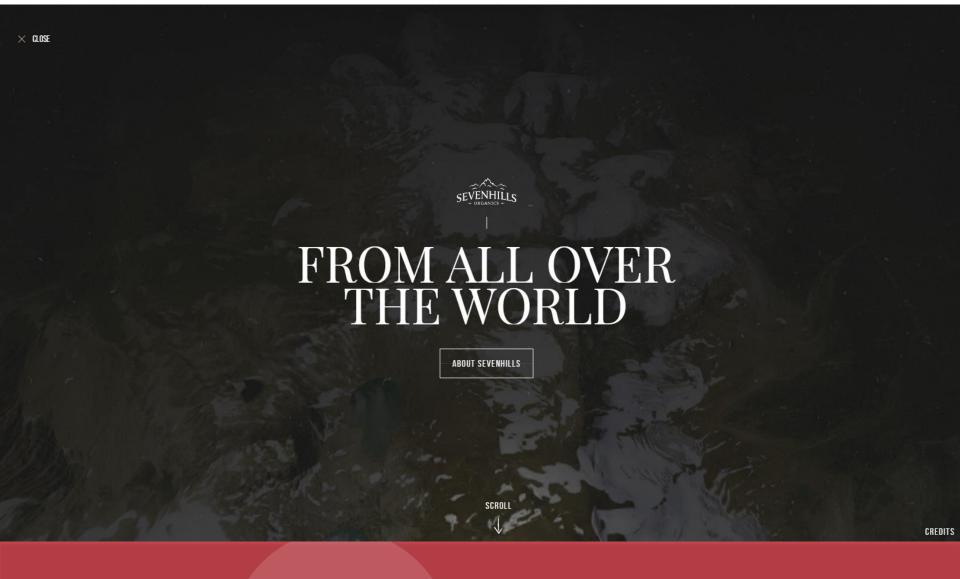
How to?



- 1. **Ecommerce tracking** Everything you want to know about how and what is being bought
- 2. Goals Focus specifically on the elements of you site that you want to push: sign up, page view and more, SM etc.
- **3. Site Search** Get to know what people think is missing from your site. What are they looking for. Get keywords.
- **4. Campaigns** ROI is the name of the game. Difficult and expensive to have others do it.
- **5. Safety net profiles** Basically backup profiles for google analytics that ensures unspoiled data.
- **6. Filter out your internal traffic** don't mess up your own data. Keep it clean, keep it valid!
- **7. Connect to Google Webmaster Tools** SEO optimization.

Best of Google analytics





Website - sevenhillswholefoods.com



1. Discoverability and Understanding:

- Can I find your way through the webpage? What is this sites mission and goal? Why does this page exist?

2. Who are they targeting, what user segment and where?

- How can you tell? What is their communication strategy?

3. What would you A/B split test?

- What surprises you? Are they breaking any rules? Do you recognize any design patterns? What do you thing the result would be?

4. Can you find examples of:

- Affordances - Signifiers

- Animations

- Mapping

- Feedback

Data visualisation

Website



Questionnaires are used in surveys.

Questionnaires are a series of questions and other prompts for the purpose of gathering information from respondents.

Can be descriptive or analytical – almost always hypothesis testing.



"This is interesting, 70% of the respondents to our survey said they don't respond to surveys."

Surveys



Quickly provides lots of **user** generated data with some reliability. (not user focused)

Well tested and recognizable to many users, which makes it easy to impose.

Takes time to make a good one, but ones done it is fairly easy to collect more data.

Data is often numerical but it still requires some analysis and the same amount of interpretation.

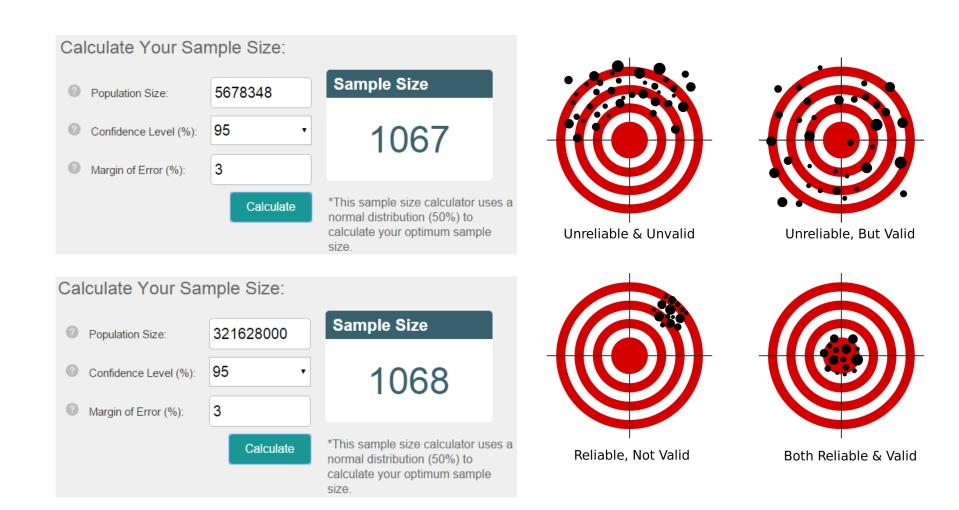
Statistical reliability and **validity** is important and difficult.





Why questionnaires

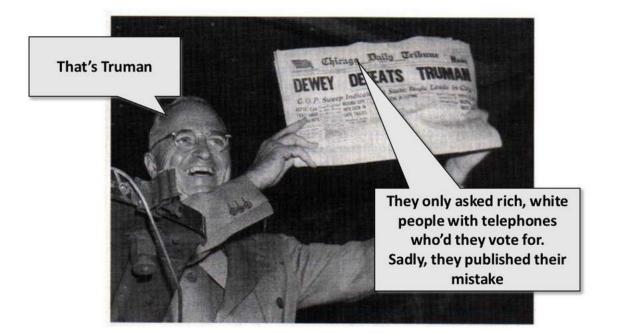




Statistical reliability



A famous sampling mistake



Something similar happened in England in spring 2015

The poll said: 36% vs. 35%

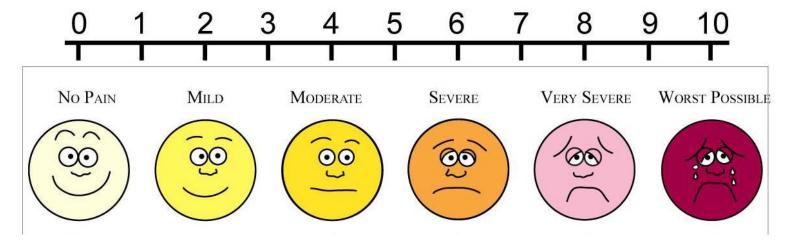
The result: 36,9% vs. 30.5

1948 presidential election



Universal Pain Assessment Tool

This pain assessment tool is intended to help patient care providers access pain according to individual patient needs. Explain and use 0-10 Scale for patient self-assessment. Use the faces or behavorial observations to interpret experssed pain when patient cannot communicate his/her pain intensity.



Wong-Baker

Likert scale (dev. 1932):

Goal: Determine attitude towards subject:

Step 1: make a number of statements.

Step 2: Judge each statement as favourable or unfavourable towards the subject. The more judges the better.

Step 3: Arrange the statements with a new scaling. Get respondents to evaluate their agreement with each statement.

Subject is: Design as a course						
	Very favourable	favourable	Neither	Un- favourable	Very un- favourable	
Design is no good	1	2	3	4	5	

	Strongly agree	Agree	Neither	Disagree	Strongly disagree
Design is no good	1	2	3	4	5

""Remeber: reverse score the negative question!!""



Likert scale (dev. 1932):

Problems and limits:

- All statements, with which a participant agrees upon, are agreed upon equally. Which point do you agree with the most?
- Dealing with mixed or complex attitudes is difficult: "Math is an important skill for computer programming, but of less use in politics."
- Attitude scales are of limited validity. They don't predict behavior very well. Words on a computer screen bear little resemblance to actual situations.



Thurstone scale (dev. 1928):

Goal: Determine attitude towards subject:

Step 1: make a number of statements.

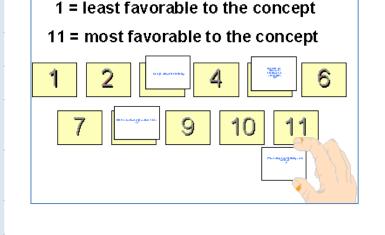
Step 2: Judge each statement as positive or negative towards the subject and assign a value (1-11). The more judges the better.

Step 3: Get respondents to pick all the statements they agree upon.

Step 4: Each participant is assigned an average score based on the value assigned to each statement they chose.

Please check all those statements with which you agree.

1. I don't approve of something that puts you out of a



- 9. It is perfectly healthy and should be legalized. (10.0)
- 10. Its use by an individual could be the beginning of a sad situation. (4.1)



Semantic Differential Scale:

Measures the meaning or relatedness of things and concepts. The connotative meaning – what is implied.

Names such as "Bakken Bears" or "Aalborg Pirates" denote animals and pirates. Their connotation is power. In contrast, the "Ulvestuen" project a different image or connotative meaning.

Rate the University park on the following dimensions:						
Safe						Dangerous
Dirty						Clean
Quiet						Noisy

Drawback: Requires respondents that are intelligent and cooperative with a good knowledge of language and willing to make fine distinctions.

















Descriptive/semantic:

-smashing

-nice

-SO**-**SO

-poor

-dreadful



player traits and habits



Velkommen til *Player traits and habits*. En indsamlig af information om læring, spil, psykologi og hverdagens vaner. Det er lige meget om du spiller mange eller få spil (eller slet ikke!) - vi vil gerne høre fra alle slags mennesker. Dine svar bliver brugt til forskning og spildesign på Erhvervsakademi Aarhus og Aarhus Universitet.

Skemaet har 4 dele. Vær sød at trykke på "next", før du lukker browservinduet. Ellers når dine svar ikke frem til os. Spørgsmålene er på engelsk, for at svare bedst muligt til andre lignende redskaber. Tænk ikke for meget over hvert spørgsmål, og gæt hvis du er i tvivl.

Spørgeskemaet er en version af en international undersøgelse, så det føles ret langt. Vi er nemlig nysgerrige over mange ting! Tag en dyb indånding og tænk på noget andet for en stund, hvis du bliver træt undervejs.

Dine svar er frivillige, fortrolige og anonyme. Alle data ender i en stor "gryde" med andre studerendes svar, og bliver aldrig analyseret på individniveau. Hvis du vil vide mere om vores forskning, er du velkommen til at skrive til andreas@psy.au.dk.

A bit of background information...

Er du lærer eller studerende på Erhvervsakademiet?

Lære

Studerende

Age

http://tinyurl.com/EAA-PTH



Q13) Which of the following best describes your age group?

- a) Under 21
- b) 21 to 24
- c) 25 to 34
- d) 35 to 44
- e) 45 to 54
- f) 55 to 64
- g) 65 or older

2. How old are you?

0-15

15-25

25-30

30-35

35-40

Methodological problems



Real ex:

Measuring whether US Christians are willing to associate with Muslims

- 1. Are you willing to live in the same country as Muslims?
- 2. Are you willing to live in the same community as Muslims?
- 3. Are you willing to live in the same neighborhood as Muslims?
- 4. Are you willing to live next door to a Muslim?
- 5. Are you willing to let your child marry a Muslim?

Point: The increasing intensity shows how scales can be used strategically to explore socio-cultural distances (rich-poor etc...)

Methodological problems



- People generally dislike filling out questionnaires
- People generally avoid extremes on a scale
- People rate responses consistently if they choose "4" for the first couple of questions they tend to continue that tendency for the rest
- Sampling errors only web-users are questioned through "surveymonkey"
- People lie and exaggerate
- People will try to outguess the survey
- People will choose an answer even if they don't feel strongly about it
- Just because people like a product doesn't mean they'll use it more just because people use a product doesn't mean they like it

Methodological problems



Do a Pilot test (pre-test)

Before the final run, pretest the form to a few people. There will ALWAYS be something that needs to be corrected.

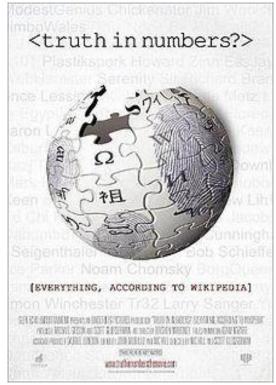
If you do not work with scales but nominal numbers the world of statistics is open! (sometimes even with scales).

In Denmark:

http://www.danmarksstatistik.dk/da/

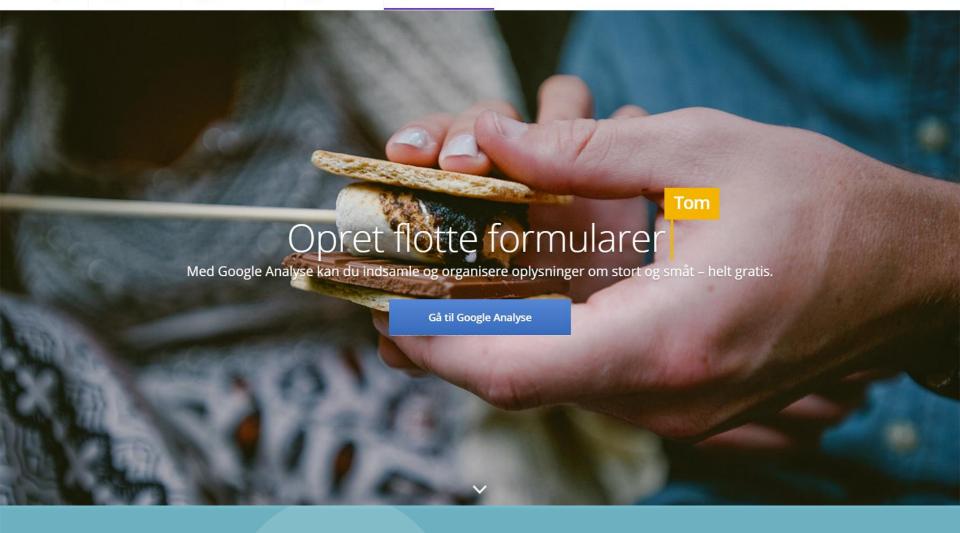
Todays tip

In small studies every person counts - so use occurrences like: "4 out of 6 did...". Median and percentage dos not make sense in these cases.



Statistics and micro studies





Google Analysis



Break 30 minutes

Jacob Nielsen's famous list of 10:

- 1. Visibility of system status
- Match between system and the real world
- 3. User control and freedom
- 4. Consistency and standards
- 5. Error prevention
- 6. Recognition rather than recall
- 7. Flexibility and efficiency of use
- 8. Aesthetic and minimalist design
- 9. Help users recognize, diagnose, and recover from errors
- 10. Help and documentation



Jacob Nielsen

A test of interface done by experts according to a set of general principles

The goal is to find and identify problems with usability, so they can be fixed.

Greek: 'Heureka' = 'I found it' Developed for software in 1990's.



- Easy and cheap yet still providing qualified insight.
- Qualitative in nature do to the subjective perspective.
- Creates a common language.

Why heuristic evaluation



"Rules of thumb" that can be used as guidelines to improve an interface.

Difficult to do alone.

Not a role belong to one person.

Verbal or written reports.

A 1-2 hour evaluation based on size.

Evaluation on: Structure – Navigation – Priority – Naming – Layout – Interaction design – Functionality – Visibility – User goals – Behaviour of site – Spelling.



How heuristic evaluation



- **1 Strive for consistency.** Consistent sequences of actions should be required in similar situations; identical terminology should be used in prompts, menus, and help screens; and consistent commands should be employed throughout.
- **2 Enable frequent users to use shortcuts.** As the frequency of use increases, so do the user's desires to reduce the number of interactions and to increase the pace of interaction. Abbreviations, function keys, hidden commands, and macro facilities are very helpful to an expert user.
- **3 Offer informative feedback.** For every operator action, there should be some system feedback. For frequent and minor actions, the response can be modest, while for infrequent and major actions, the response should be more substantial.
- **4 Design dialog to yield closure.** Sequences of actions should be organized into groups with a beginning, middle, and end. The informative feedback at the completion of a group of actions gives the operators the satisfaction of accomplishment, a sense of relief, the signal to drop contingency plans and options from their minds, and an indication that the way is clear to prepare for the next group of actions.

Ben Schniderman "Eight Golden Rules"



- **5 Offer simple error handling.** As much as possible, design the system so the user cannot make a serious error. If an error is made, the system should be able to detect the error and offer simple, comprehensible mechanisms for handling the error.
- **6 Permit easy reversal of actions.** This feature relieves anxiety, since the user knows that errors can be undone; it thus encourages exploration of unfamiliar options. The units of reversibility may be a single action, a data entry, or a complete group of actions.
- **7 Support internal locus of control.** Experienced operators strongly desire the sense that they are in charge of the system and that the system responds to their actions. Design the system to make users the initiators of actions rather than the responders.
- **8 Reduce short-term memory load.** The limitation of human information processing in short-term memory requires that displays be kept simple, multiple page displays be consolidated, window-motion frequency be reduced, and sufficient training time be allotted for codes, mnemonics, and sequences of actions.

Ben Schniderman "Eight Golden Rules"



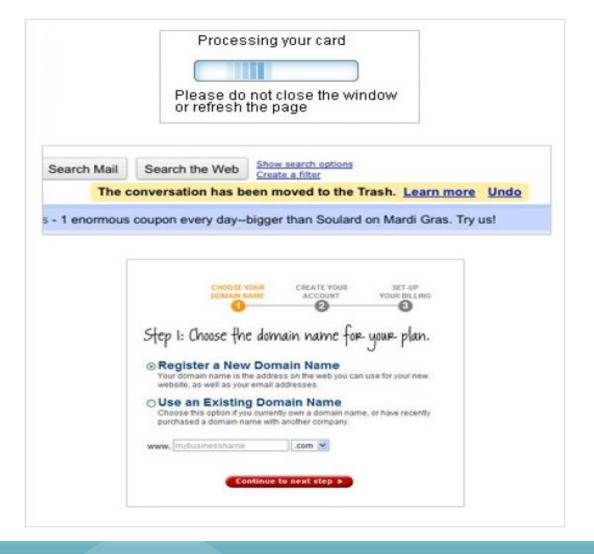
- 1. Visibility of system status
- 2. Match between system and the real world
- 3. User control and freedom
- 4. Consistency and standards
- 5. Error prevention
- 6. Recognition rather than recall
- 7. Flexibility and efficiency of use
- 8. Aesthetic and minimalist design
- 9. Help users recognize, diagnose, and recover from errors
- 10. Help and documentation



nngroup.com/articles/tenusability-heuristics/

Jacob Nielsen





1. Visibility of system status



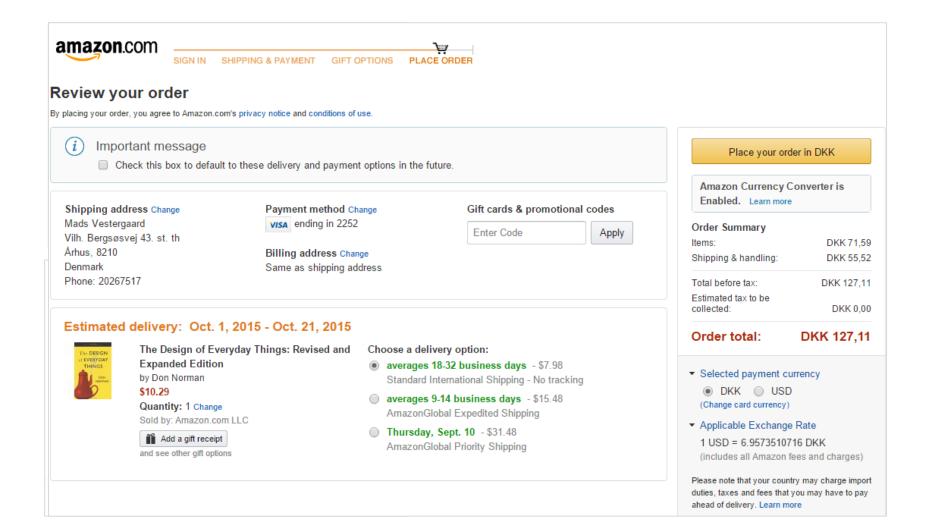
The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.



Metaphors, Mapping, Feedback, Discoverability and Understanding

2. Match between system & real world





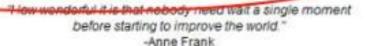
3. User control & freedom



SERVICE LEARNING & COMMUNITY ACTION :: INFO FOR COMMUNITY ORGANIZATIONS

- Service Learning Home
- Students
- Faculty.
- Community Organizations
- Service Learning News
- Additional Resources
- · FAQ

INFORMATION FOR COMMUNITY-BASED AND NON-PROFIT ORGANIZATIONS



Emerson College's communication and arts curricula prepare students to make creative and often tangible contributions to community-based organizations. These unique contributions include market research and public relations assistance, press releases, media products, theatre education, and many other arts and communication services.

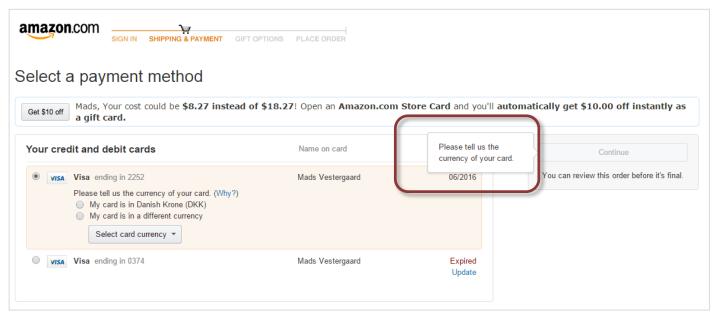


4. Consistency and standards





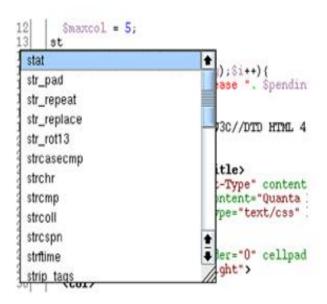




5. Error prevention







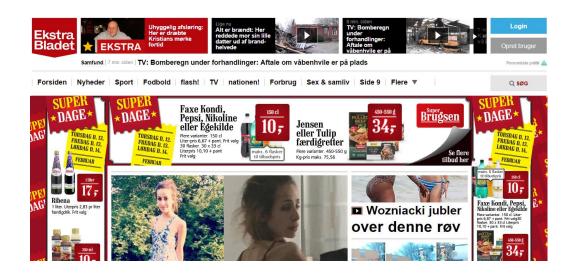
6. Recognition rather than recall



Shortcut Description	Windows	Mac
Select the entire page	Ctrl + A	Cmd + A
Copy selected information	Ctrl + C	Cmd + C
Cut selected information	Ctrl + X	Cmd + X
Paste copied information	Ctrl + V	Cmd + V
Open a file	Ctrl + O	Cmd + O
Save a file	Ctrl + S	Cmd + S
Undo the last operation	Ctrl + Z	Cmd + Z
Redo the last operation	Ctrl + Y	Shift + Cmd + Z
Find text in the current document/page	Ctrl + F	Cmd + F
Task Manager	Ctrl + Shift + Esc	Alt + Cmd + Esc

7. Flexibility and efficiency of use



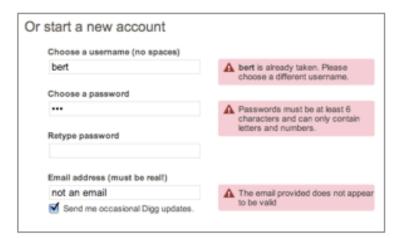


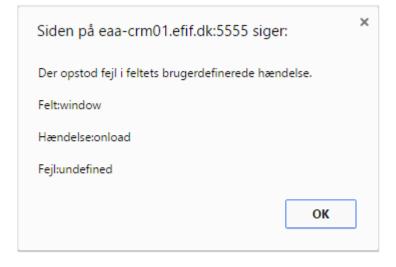


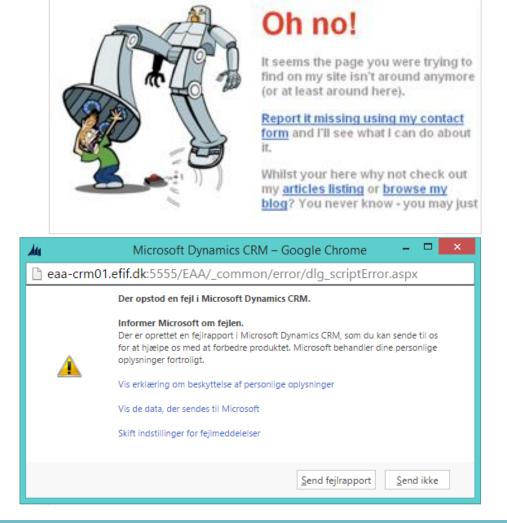
Google.dk på: Føroyskt

8. Aesthetic and minimalist design



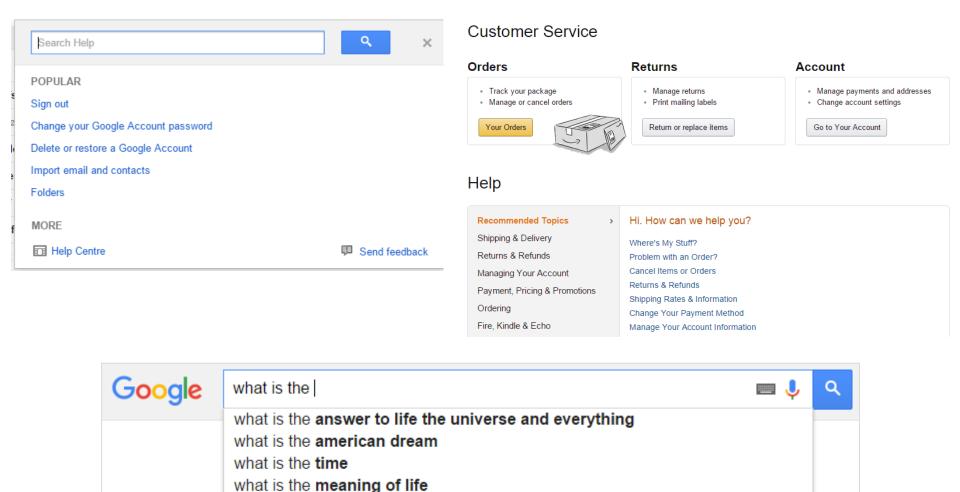






9. Error recovery





10. Help and documentation

Tryk på Enter for at søge.

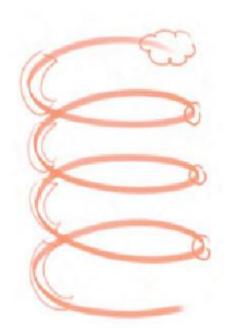


Heuristic testing is just one step:

Based on the results of the heuristic testing, the next iteration of the design situation can begin. Heuristics can help to inform and direct scenarios

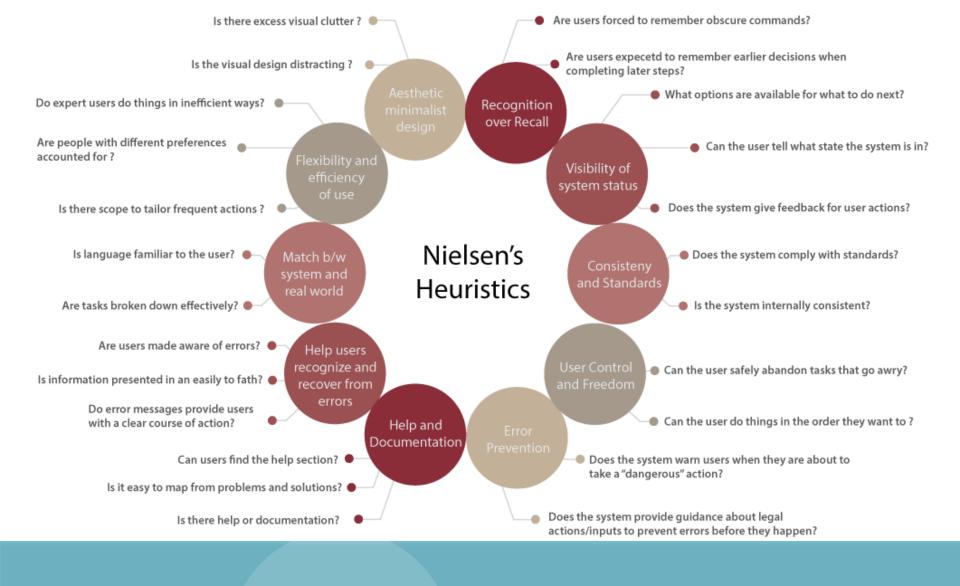
No heuristic method predicts end-user problems better than actual usability testing! In short:

- 1. Do an expert review by evaluating with heuristics
- 2. Clean the interface yourself
- 3. Identify goals for a usability test
- 4. Do a "real" usability study



Heuristic testing





Heuristic Analysis/Expert Review

In your groups (from the database class) evaluate an existing website in relation to Jakob Nielsen's 10 heuristics for good usability.

Part 1 – Individually:

Use the principles to discover problems and improve usability.

Part 2 – In groups:

Collect the identified problems and give suggestions on how to improve.

You should end up with: A written report documenting:

- 1. Which heuristics are "violated"
- 2. How severe is the issue (on a scale of "fix now" to "can wait").
- 3. Ideas for a solution.

Evaluate - bmw.com/com/en/



... just sit down with your users and the interface. **Watch** what they do and **listen** to them to think out loud.

- + You evaluate your design decisions
- + One step closer to understanding users
- + Discover opportunities unknown
- + Can be made cheap and quick
- It will affect how they use the interface
- It is not a true picture
- Time consuming
- Results might be hard generalize and act upon



Site down usability test - See Unger



How to cope?

- 1. Define your goals what do you want to test build an hypothesis to guide you.
- 2. Formulate fitting tasks and instruct the user to solve the. Don't tell them how!
- 3. Don't follow your guide. Let the user make the choices observe only.
- 4. Look both at the interaction and the users reaction. Listen carefully to what they say.
- 5. Evaluate your observations against you hypothesis. What did you expect and what actually happened. Why the difference?



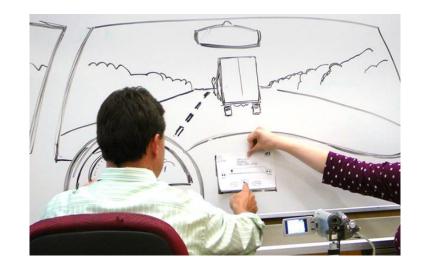
Using heuristics (or similar) or planning a workshops can help. Plus use video and tape recorder.

Site down usability test

Cooper et.al

Choosing between research or testing – research gives the better result.

More important to spend time making considered design decisions based on a solid research foundation than to test half-baked design

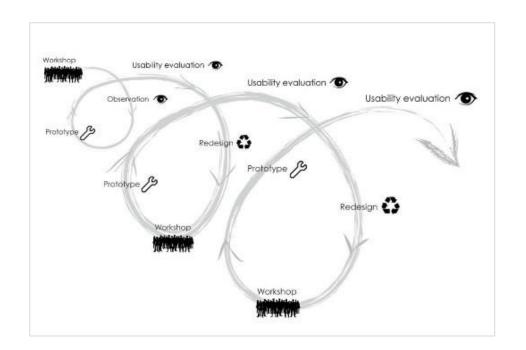


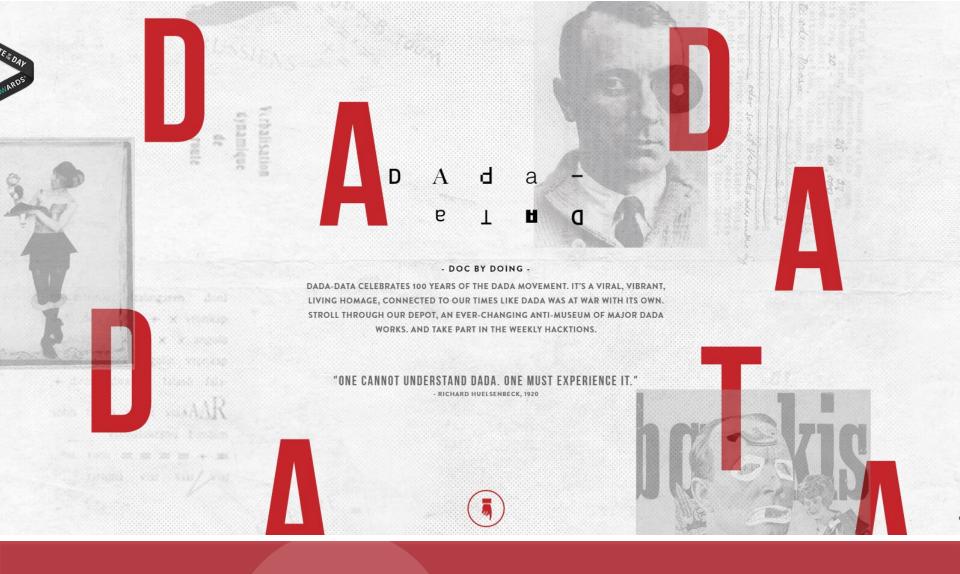
What do you thing? User test



Russ Unger et.al

The primary purpose of "concept exploration" is to understand the kinds of responses and ideas that are elicited from your user groups when faced with a set of design elements.





Website



Visibility of system status: The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.

Match between system and the real world: The system should speak the users' language, with words and concepts familiar, rather than system-oriented terms.

User control and freedom: Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave

Consistency and standards: Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow <u>platform conventions</u>. Or actions

Error prevention: Even better than good error messages is a careful design which prevents a problem from occurring in the first place

Recognition rather than recall: Minimize the user's memory load by making objects, actions, and options visible.

Flexibility and efficiency of use: Accelerators -- unseen by the novice user.

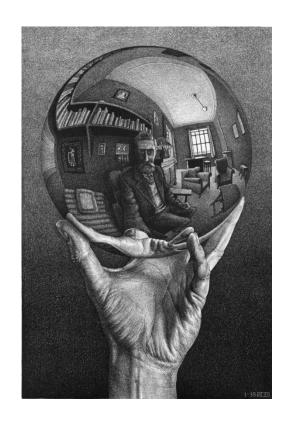
Aesthetic and minimalist design: Dialogues should not contain information which is irrelevant or rarely needed.

Help users recognize, diagnose, and recover from errors: Error messages should be expressed in plain language, precisely indicate the problem, and solution.

Help and documentation: Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation.

Evaluating designs

- Quantitative research.
- A/B Split testing to increase conversion.
- Eye tracking and Heat map to know behaviour.
- Questionnaires to get to know attitudes and associated meanings.
- Jakob Nielsen's Heuristics.
- User tests.



What happened today?



Next time:

Topic: Responsive and adaptive web development. Native apps and Mobile first, plus basic and advanced navigation. How to design Forms.

Read / Do:

Check the examples in (you will be working with it): [Frost] Text 1 - On mobile navigation

[Frost] Text 2 - On complex mobile navigation

[Cooper] About Postures for web and mobile p. 218 - 230 [Wroblewski] A reason for mobile first p. 7 - 17 [Krug] Mobile; not just a city in Alabama p. 143 - 163 (easy)

Next time

