

Evaluating Designs

Other Approaches to Evaluate
Designs
Final Elements

Evaluating Designs



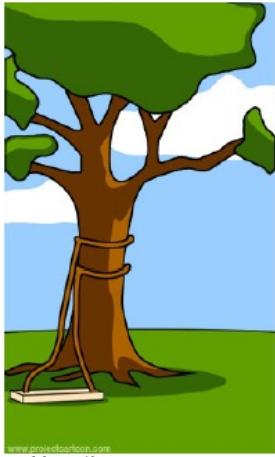
How the customer explained it



How the project leader understood it



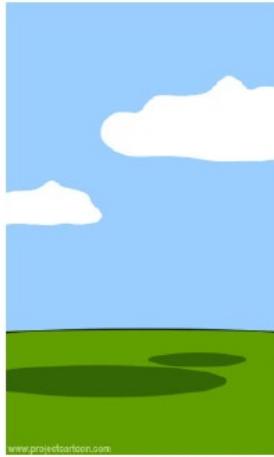
How the analyst designed it



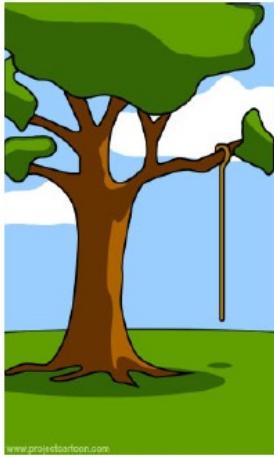
How the programmer wrote it



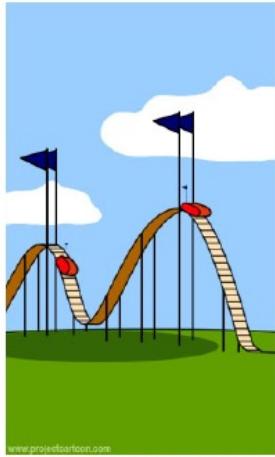
How the business consultant described it



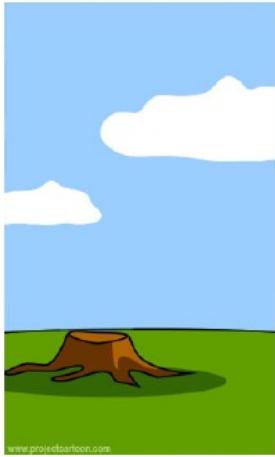
How the project was documented



What operations installed



How the customer was billed



How it was supported



What the customer really needed

Why, what, where, when and how to evaluate

Iterative design & evaluation is a continuous process that examines: Why, what, where and when...

- Why: to check users' requirements and that users can use the product and they like it.
- What: a conceptual model, early prototypes of a new system and later, more complete prototypes.
- Where: in natural and laboratory settings.
- When: throughout design; finished products can be evaluated to collect information to inform new products.
- How: with users or without real users

How to Evaluate

Two ways to evaluate designs

- With real users (usability studies with participants)
- Without real users (experts, analysis, inspections, etc.)

Evaluations with Users	Evaluations without Users
Controlled experiment	Cognitive Walkthroughs
Field studies	Heuristics
Interviews	Key-stroke
Questionnaires/surveys	Simulation
Observation	



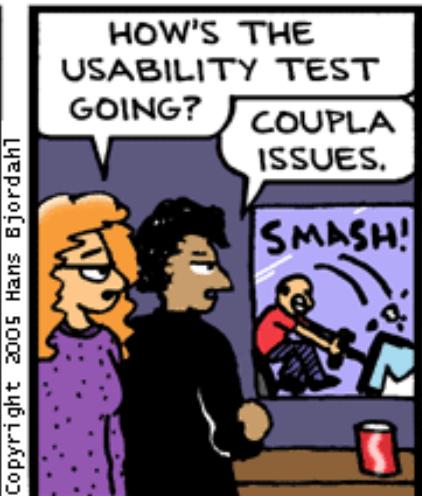
Bug Bash by Hans Bjordahl



THAT'S SO STUPID!
I CAN'T BELIEVE
IT!



ARRRGH!
SMASH!



HOW'S THE
USABILITY TEST
GOING?
COUPLA
ISSUES.

Copyright 2005 Hans Bjordahl

<http://www.bugbash.net/>

Types of evaluation



Controlled settings involving users, e.g., usability testing & experiments in laboratories



Natural settings involving users, e.g., field studies to see how the product is used in the real world.



Any settings not involving users, e.g., consultants' critique, usability experts; to predict, analyze & model/simulate aspects of the interface analytics.

For example: Inspections

- Experts use their knowledge of users and technology to review usability
- Expert critiques can be formal or informal reports.
 - *Walkthroughs* involve stepping through a pre-planned scenario noting potential problems.
 - *Heuristic evaluation* is a review guided by a set of heuristics.



Cognitive Walkthroughs

- Good for generating lots of design ideas from different backgrounds and perspectives
- Representative tasks selected for the cognitive walk-through by the moderator
 - Each task is broken down into a series of steps along with a description of users, their experience, and prior knowledge (e.g., from a persona)
- Expert is told the assumptions about user population (persona/s), context of use (task scenarios), task details (use-cases).
- One or more experts walk through the design prototype with the scenario
- Experts are **guided by 3 questions:**
 - **Will the user know what to do?**
 - **Will the user know how to do it?**
 - **Will the user know they made the right action/selection?**

Advantages of Cognitive Walkthroughs

- Can be an effective inspection method not just by cognitive scientists and usability specialists but also by novice evaluators
- Helpful in picking out interface problems at an early stage, and works particularly well together with a user-centred design approach and the development of user personas

Disadvantages of Cog Walkthroughs

- Some issues with cognitive walkthroughs:
 - Answering the three questions and discussing the answers can take a long time
 - Designers can be defensive often invoking cognitive theory to explain design
- Another approach – reduce the questions (but more of a learning curve):
 1. Will the user know what to do at this step?
 2. If the users do the right thing, will they know that they did the right thing and are making progress to towards their goal?

Variation of Walkthrough – Pluralistic Walkthrough

- Users, developers and usability experts work together through a task scenario and discuss usability issues
- Each group of experts is asked to assume the role of typical users
- Once the team has all written down their actions, they then discuss the actions making comments on issues and suggestions (converge on decisions)
- Can do this between tasks or at the end.

Adv/Disadv of Pluralistic Approach

- Advantages of Pluralistic approach:
 - Strong focus on users at task detailed levels (e.g., looking at steps)
 - Group brings variety of opinions (maybe expertise) for interpreting each step of interaction
- Limitations of approach
 - Getting experts together all at once
 - Only limited number of scenarios and paths through the interface can be explored because of time constraints

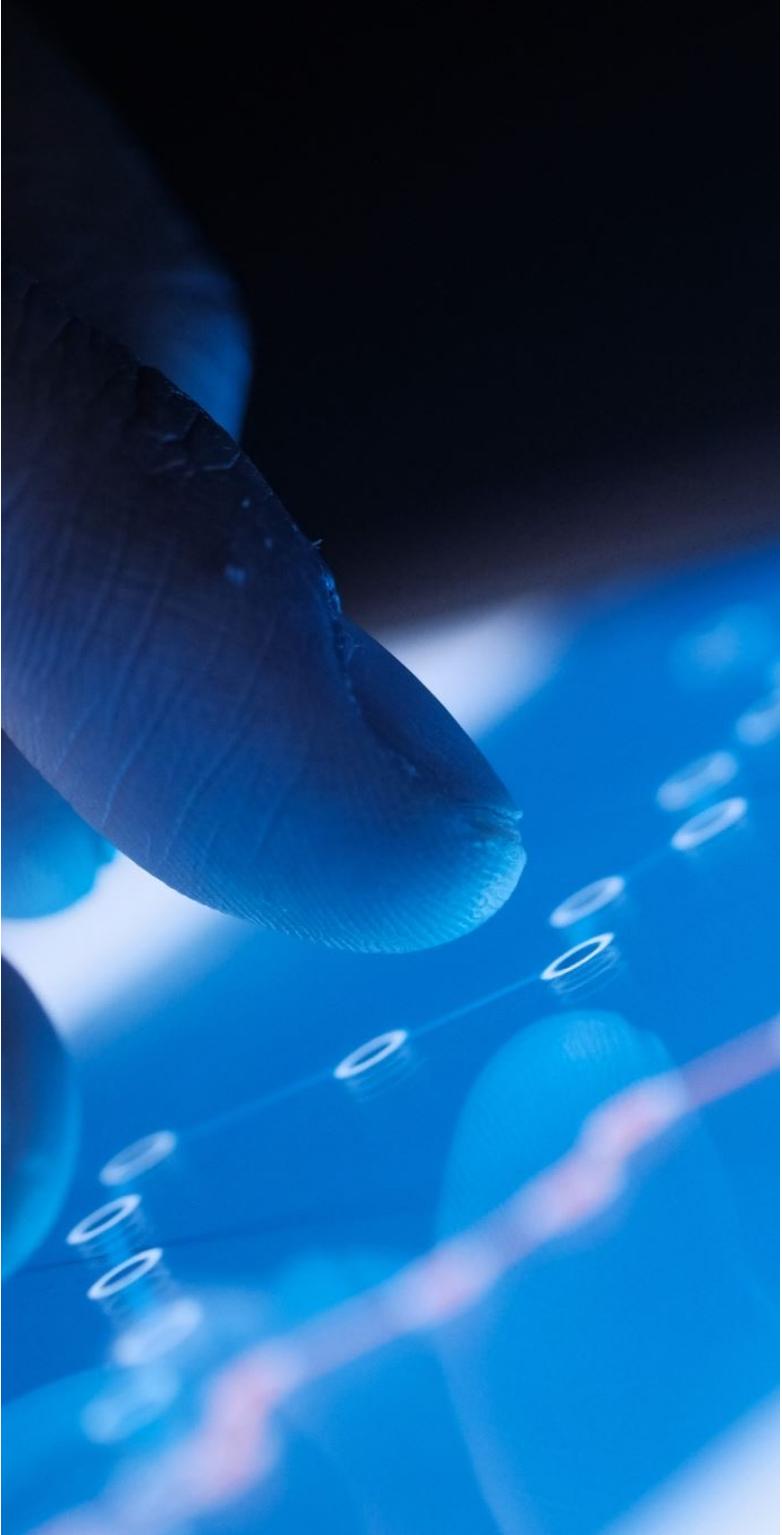


Another Inspector Approach: Heuristics

- First developed by Jakob Nielsen and his colleagues in the early 90's
 - Experts are guided by set of usability principles (heuristics) evaluate user interface elements (eg., dialog boxes, navigation structure, menus, messages, etc.) based on these guidelines
 - The heuristics closely resemble the *high-level design principles* and cognitive background that we looked at (e.g., consistency, reducing memory load, etc)

You should check out Nielsen's website:

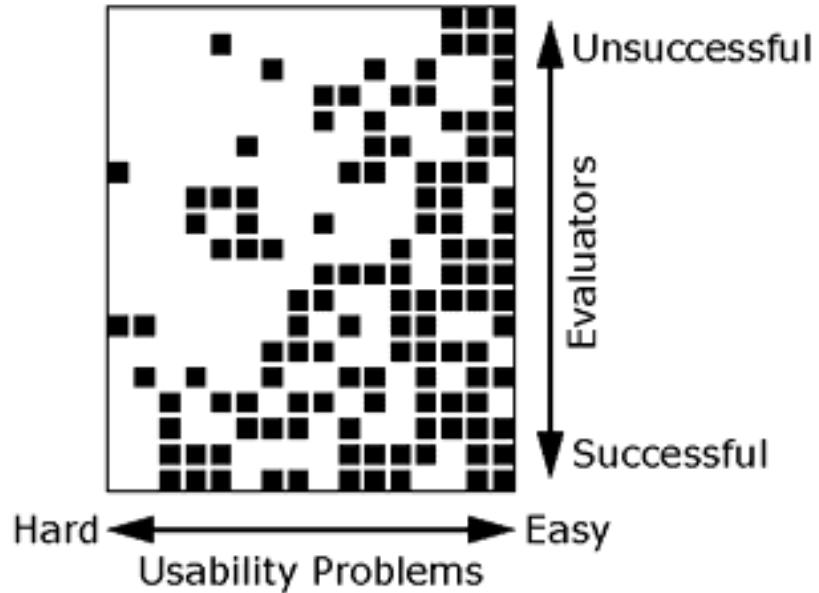
http://www.useit.com/papers/heuristic/heuristic_list.html



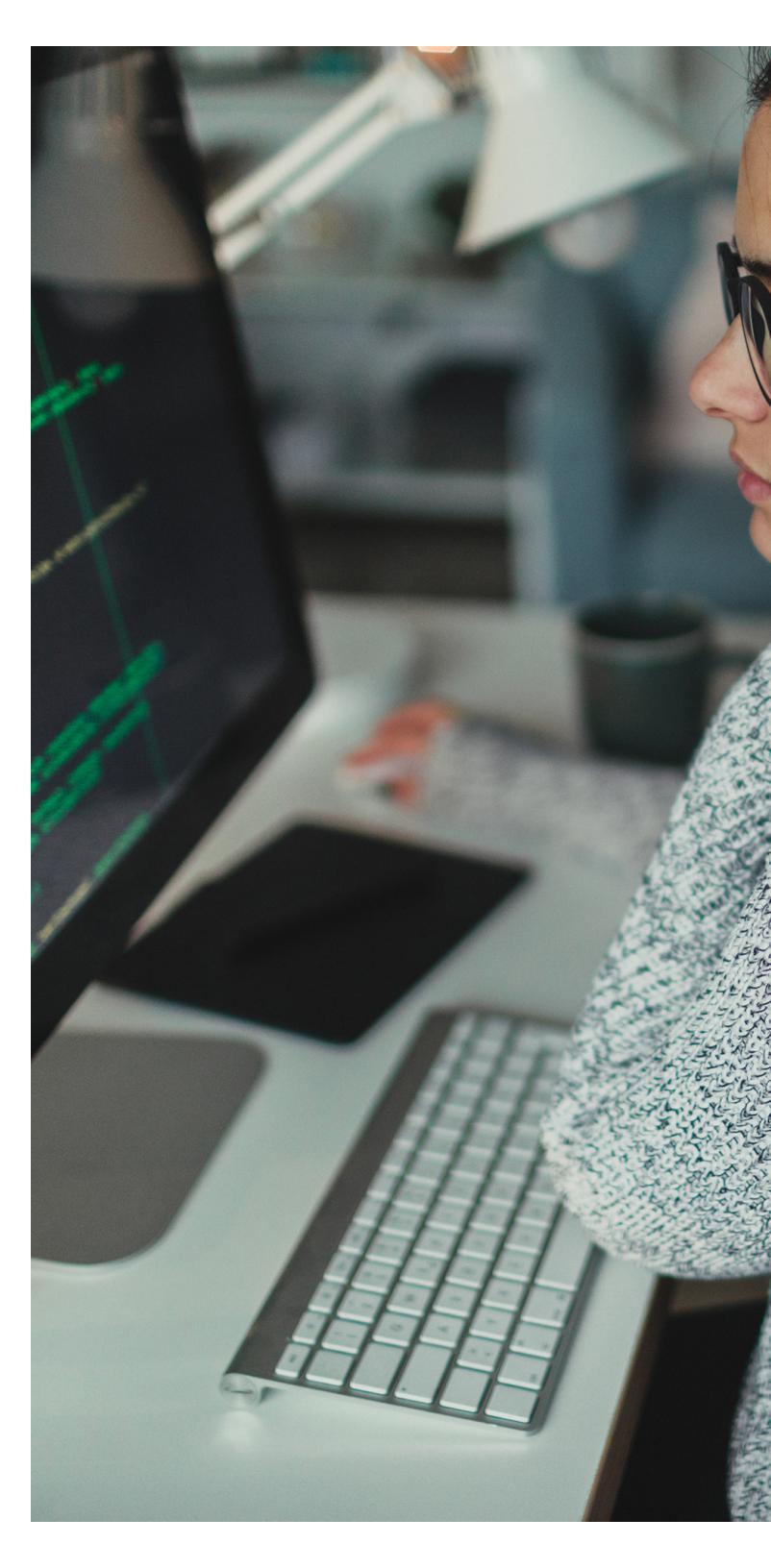
Heuristics

- Experts judge the interface using the heuristics for different aspects of the interface (can find **up to 75%** of usability problems)
- E.g., if you are evaluating a new email system, you might be concerned in particular with the last heuristic (to determine if a new user could use help to figure how to create a new email and send it)
- Each evaluator would go through the interface several times inspecting the different interface elements and comparing them with the heuristics
- Evaluators come together afterwards to discuss the issues and rate them

Heuristics

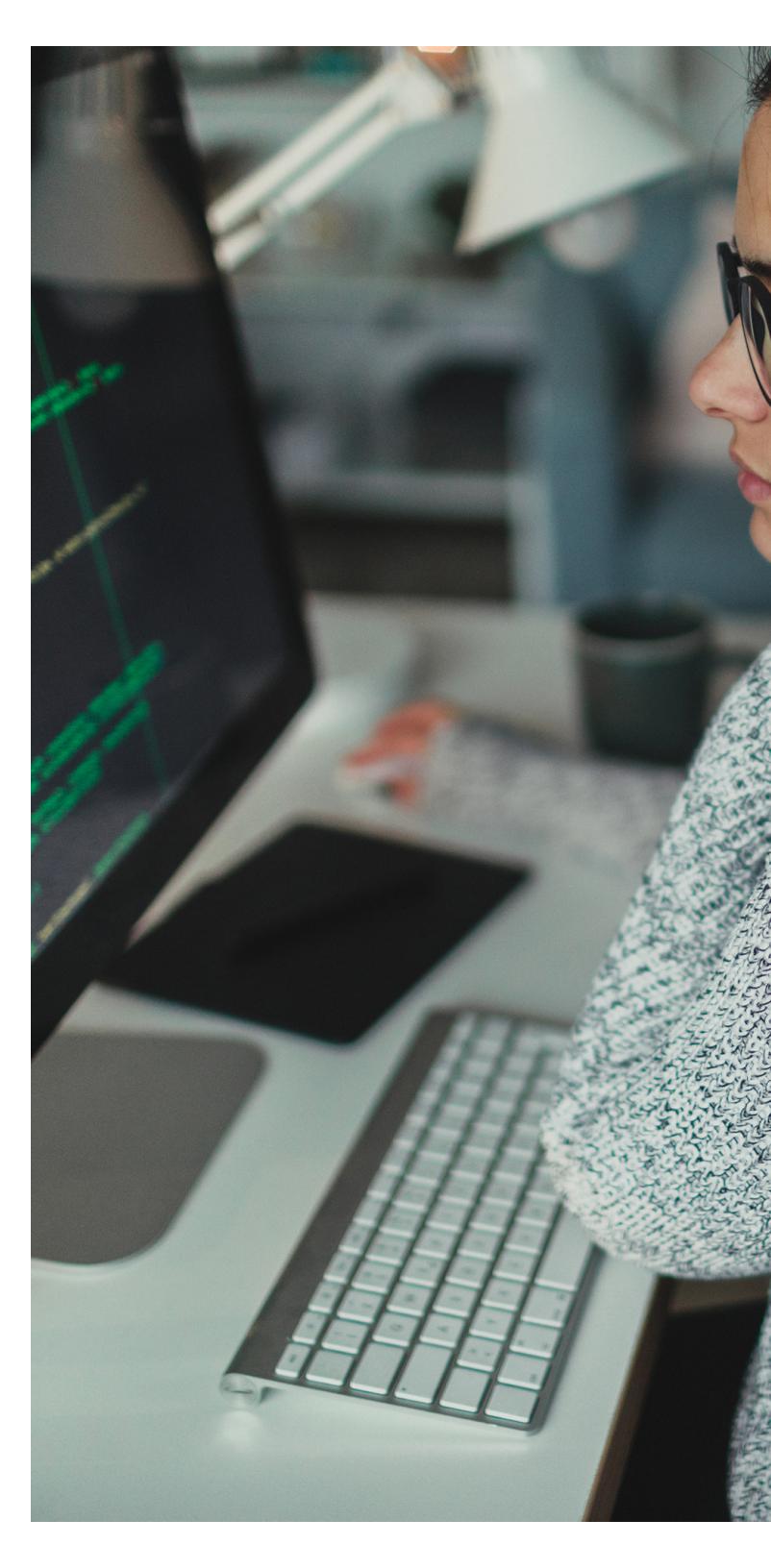


- How many evaluators?
 - One can catch many but not all (e.g., one evaluator may focus on different usability issues than someone else)
 - **Recommended number 3-5** (easy for all to find the main issues, need more than a few to find the more obscure ones) but often depends on the complexity of the interface and other factors (e.g., skill needed, tasks to be done, etc.)
 - 1 catches ~35%
 - **5 catch 75%**
 - Also – multiple experts can be costly
 - Early on – even one or two evaluators can identify most problems in a working prototype

A photograph showing a person from the side, wearing glasses and a patterned shirt, working at a computer keyboard. The monitor in front of them displays some green and white graphical data. A small cup sits on the desk next to the keyboard.

Problems found with a Single Evaluator

- Average over six case studies
 - 35% of all usability problems;
 - 42% of the major problems
 - 32% of the minor problems
- Not great, but
 - finding some problems with one evaluator is much better than finding no problems with no evaluators!

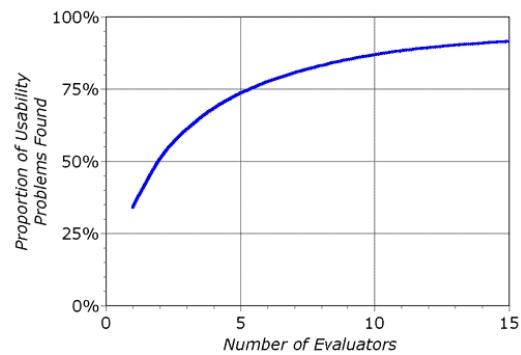


Problems found with a Single Evaluator

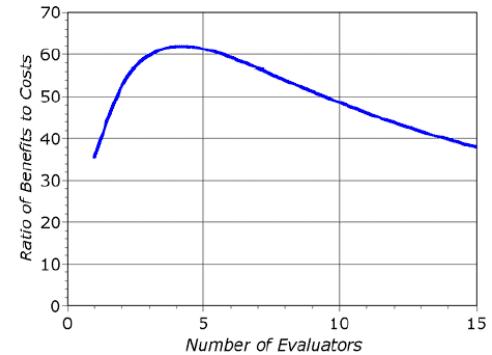
- Varies according to
 - difficulty of the interface being evaluated
 - the expertise of the inspectors
- Average problems found by:
 - novice evaluators - 22%
 - no usability expertise
 - regular specialists - 41%
 - expertise in usability
 - double specialists - 60%
 - experience in both usability and the particular kind of interface being evaluated
 - also find domain-related problems
- Tradeoff: novices poorer, but cheaper!

Decreasing Returns

problems found



benefits / cost



basic idea: ask 3-5 usability experts to inspect your system to identify as many problems as possible – called “discount usability”



Nielsen's heuristics:

1. Visibility of system status (Feedback)

- System should always inform users about what is going on, through appropriate feedback in a reasonable time

What are some examples?

- Does every display begin with a title or header that describes screen contents?
- Is there a consistent icon design scheme and stylistic treatment across the system?
- After the user completes an action (or group of actions), does the feedback indicate that the next group of actions can be started?
- Is there visual feedback in menus or dialog boxes about which choices are selectable?

Upload the file



Type new password: Six-characters minimum; case sensitive

Password strength:

Messages

Nielsen's heuristics:

1. Visibility of system status (Feedback)

- System should always inform users about what is going on, through appropriate feedback in a reasonable time

What are some examples?

- Does every display begin with a title or header that describes screen contents?
- Is there a consistent icon design scheme and stylistic treatment across the system?
- After the user completes an action (or group of actions), does the feedback indicate that the next group of actions can be started?
- Is there visual feedback in menus or dialog boxes about which choices are selectable?

Upload the file

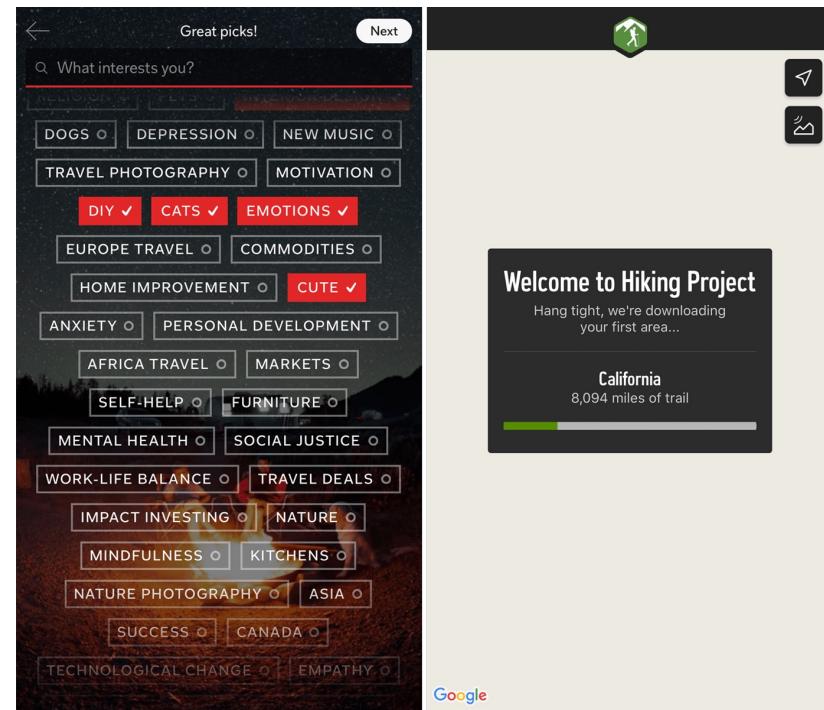


Type new password: Six-characters minimum; case sensitive

Password strength:

Messages

Feedback (more examples)



What about systems that don't have a visual interface (e.g., Siri or Google home, Amazon)?
What about keyboards on mobile phones?



Nielsen's heuristics:

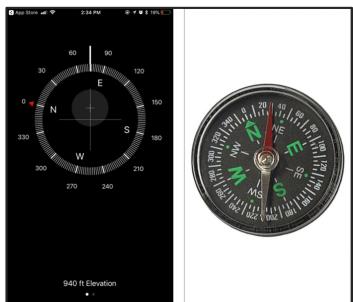
2. Match between system and the real world (Metaphor)

- System should speak the user's language (words, phrases, concepts that are familiar with the user)
- Make information appear in a natural and logical order
- Examples:
 - Are icons concrete and familiar?
 - Are menu choices ordered in the most logical way, given the user, the item names, and the task variables?
 - If there is a natural sequence to menu choices, has it been used?



iTunes – icons, logical order

Match between system and real world (more examples)



PREVENTIVE CARE IS IMPORTANT TO YOUR HEALTH

Health Benefits & Coverage

WHAT IS PREVENTIVE CARE?

Preventive care (sometimes called "preventive care") is routine health care that includes screenings, services and counseling to help prevent illness, disease or other health problems. It's covered by your plan because it can help you stay healthy and lets you know if you may develop a health problem.

Don't let disease sneak up on you. Preventive care could save your life.

UNDERSTANDING YOUR PREVENTIVE CARE BENEFITS

As mentioned in your benefits booklet, many screenings and tests are covered at 100% with no out-of-pocket costs to you as part of your health plan. Use your preventive care and stop health problems before they get serious.

[SUMMARY OF PREVENTIVE CARE SERVICES](#)

WHICH TYPES OF SERVICES ARE COVERED AT 100%?

Preventive care that is 100% covered includes many important services like the ones listed below. This list is not complete, so please see the full list of preventive care services and any limitations on the [Men](#), [Women](#) and [Children](#) pages.

ABACUSNEXT

Private Cloud ▾ Business Software ▾ Accounting Solutions ▾ Legal Solutions ▾ Success Stories ▾ Our Company ▾ Partners ▾ More ▾

SEARCH **Contact Us** **Get a Free Demo**

Abacus Private Cloud Quick Facts

- ✓ 8 Data Centers
- ✓ 100% Renewable Data Center Power
- ✓ SOC1/SOC2/SOC3/SSAE16 Certifications
- ✓ 256-bit AES Data Encryption
- ✓ 99.999% Uptime Service Level Agreement
- ✓ Unlimited Cloud Platform Support
- ✓ No Fee for After Hours Support
- ✓ No Fee for Client Data Export
- ✓ No Fee for Device Reconfiguration
- ✓ Accessible Anywhere, Anytime
- ✓ Up to 60 Months Price Guarantee

Law Firm Business Continuity Management

Abacus Private Cloud is complete with robust security and compliance policies, including two-factor authentication, five tiers of physical, network and data security, information handling and data privacy protocols in addition to custom Business Continuity solutions that address both catastrophic and clinical events to protect your firm's confidential and irreplaceable case and client data from the unpredictable.

[Learn More](#)

Full-Spectrum Security

Abacus Private Cloud is fully integrated backup and disaster recovery cloud technology in addition to custom Business Continuity solutions that address both catastrophic and clinical events to protect your firm's confidential and irreplaceable case and client data from the unpredictable.

[Learn More](#)

Zero Capital Investment

Private Cloud environments reduce upfront capital expenses such as server hardware and software, backup hardware and software, network infrastructure and installation fees. Leverage your operating expenses budget while speeding time to deployment with Abacus Private Cloud virtualization technology.

Complete Ownership of Data

Unlike other Public Cloud storage and SaaS providers that may keep you from accessing your own data, with Abacus Private Cloud you own your data. Upon termination of any business relationship, you will always have access to your irreplaceable case and client data.

"BlueCross BlueShield of North Carolina provides clear answers to questions, uses familiar language, and includes references that are easy to understand."

"The About page for Abacus legal software's website uses obscure, overly technical language that places much burden on the user to connect the words with their meaning."

Nielsen's heuristics:

3. User control and freedom (Navigation)

- Users often make mistakes and will need a clearly marked ‘exist’ – support undo and redo

Examples:

- Can users cancel out of operations in progress?
- Are character edits allowed in data entry fields?

	A	B	C	D
1	Item	Quantity	Price	Total
2	Tacos	40	\$5.00	=B2 * C2
3				

Shows cell and column ids, the action and can accept or delete action

The screenshot shows the Wufoo Form Gallery interface. At the top, there are navigation links: Home → Gallery → Templates. Below this is the 'FORM GALLERY' logo. To the right, there are two buttons: 'Download HTML' and 'Add to Wufoo'. A sidebar on the left lists various template categories with numbered options: 1 Customer Satisfaction Survey, 2 Cancellation Survey, 3 Business Demographic Survey, 4 Web Site Visitor Survey, 5 Tech Support Satisfaction Survey, and 6 Health Survey. The main content area displays the 'Customer Satisfaction Survey' template. It includes a header with the Wufoo logo, a brief instruction 'Please take a few moments to complete this satisfaction su...', and a question 'How long have you used our product / service?'. There are four radio button options: 'Less than a month', '1-6 months', '1-3 years', and 'Over 3 Years'.

Wufoo - clearly shows where they are, and where they can go (menu)

Nielsen's heuristics:

4. Consistency and standards (Consistency)

- Users shouldn't wonder whether different words, images, situations, layouts or actions mean the same thing.
- Follow conventions

Examples:

- Are icons labeled?
- Does the menu structure match the task structure?
- Are system objects named consistently across all prompts in the system?
- Are user actions named consistently across all prompts in the system?



Gmail - email

Nielsen's heuristics:

5. Error Prevention (Prevention)

- Even better than a good error message – prevent the problem from occurring (test, test...)

Examples:

- Have dots or underscores been used to indicate field length?
- Are data inputs case-blind whenever possible?
- Do data entry screens and dialog boxes indicate the number of character spaces available in a field?



Filling in form

design		Advanced Search
design within reach	5,350,000 results	Preferences
designer handbags	3,430,000 results	Language Tools
designer shoes	2,630,000 results	
designer clothes	3,120,000 results	
designer dresses	1,110,000 results	
design sponge	9,930,000 results	
designer	285,000,000 results	
design museum	13,600,000 results	
designers guild	530,000 results	
designer jeans	2,010,000 results	
	close	

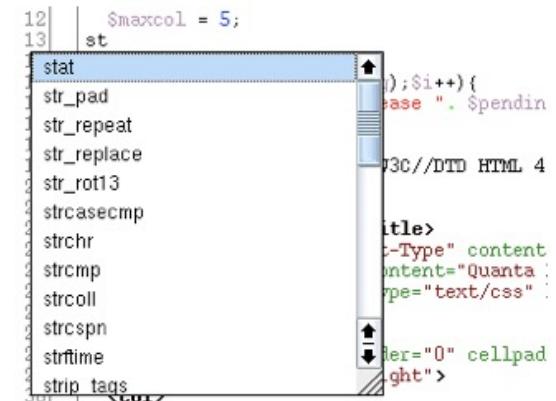
Nielsen's heuristics:

6. Recognition rather recall (Memory)

- Minimize the user's memory load by making objects, actions, and options visible
- Users shouldn't have to remember info from one part of the system to another
- Instructions for use should be visible or easily retrievable

Examples:

- For question-and-answer interfaces, are visual cues and white space used to distinguish questions, prompts, instructions, and user input?
- Have items been grouped into logical zones, and have headings been used to distinguish between zones?
- Are inactive menu items grayed out or omitted?
- Is color coding consistent throughout the system?
- Has the same color been used to group related elements?



Type Ahead for Coding

Nielsen's heuristics:

7. Flexibility and efficiency of use (Efficiency)

- Allow users to tailor frequent actions but provide easy actions for novice users

Examples:

- If the system supports both novice and expert users, are multiple levels of error message detail available?
- Can expert users bypass nested dialog boxes with either type-ahead, user-defined macros, or keyboard shortcuts?

Common Shortcuts	
Add Action	Return
New Window	⌘N
Synchronize with Server	⌃⌘S
Clean Up	⌘K
Planning Mode	⌘1
Context Mode	⌘2
Inbox	⌃⌘1
Quick Entry	⌃⌃Space
<small>Quick Entry's shortcut can be customized in Preferences</small>	

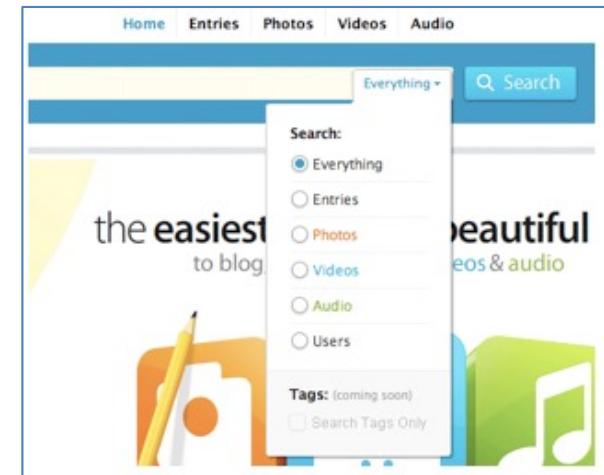
Nielsen's heuristics:

8. Aesthetic and minimalist design (Design)

- Dialogues should contain relevant information
- Layout should follow basic design principles (e.g., structure, simplicity, consistency, tolerance)

Examples:

- Is only (and all) information essential to decision making displayed on the screen?
- Are all icons in a set visually and conceptually distinct?
- Are menu titles brief, yet long enough to communicate?



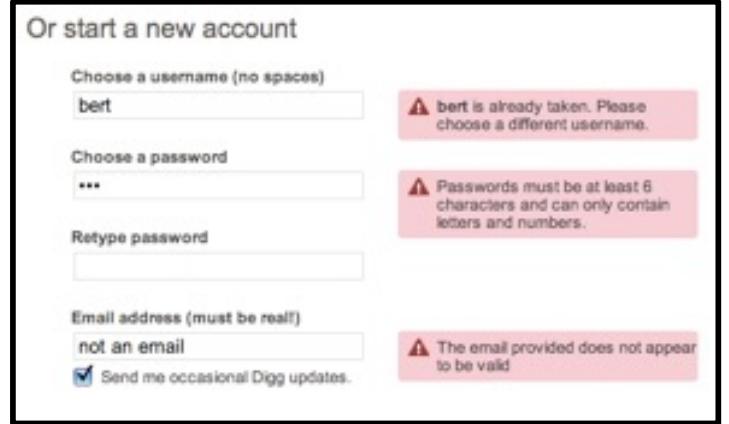
Nielsen's heuristics:

9. Help users recognize, diagnose, and recover from errors (Recovery)

- Error messages should be in plain language that id the problems and suggests a solution

Examples:

- Are prompts stated constructively, without overt or implied criticism of the user?
- Are error messages worded so that the system, not the user, takes the blame?
- Do error messages suggest the cause of the problem?



Or start a new account

Choose a username (no spaces)
bert ⚠ bert is already taken. Please choose a different username.

Choose a password
*** ⚠ Passwords must be at least 6 characters and can only contain letters and numbers.

Retype password

Email address (must be real!)
not an email ⚠ The email provided does not appear to be valid.

Send me occasional Digg updates.

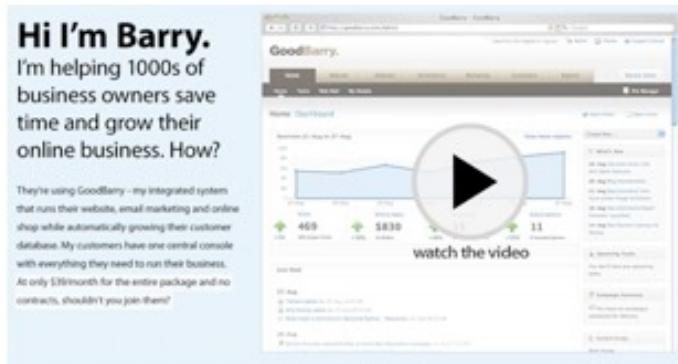
Nielsen's heuristics:

10. Help and documentation

- Even though systems should be intuitive to use and not be dependent on documentation, it may still be necessary to have help and documentation
- Should be easy to search, focus on user tasks, list concrete steps to be carried out and not be too long

Examples:

- Are on-line instructions visually distinct?
- Is the help function visible; for example, a key labeled HELP or a special menu?
- If menu items are ambiguous, does the system provide additional explanatory information when an item is selected?
- Navigation: Is information easy to find?



Heuristics

- Experts judge the interface using the heuristics for different aspects of the interface (can find up to 75% of usability problems)
- E.g., if you are evaluating a new email system, you might be concerned in particular with the last heuristic (to determine if a new user could use help to figure how to create a new email and send it)
- Each evaluator would go through the interface several times inspecting the different interface elements and comparing them with the heuristics
- Evaluators come together afterwards to discuss the issues and rate them

Excerpt from a Heuristics Worksheet

Heuristic Evaluation - A System Checklist

1. Visibility of System Status

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

#	Review Checklist	Yes	No	N/A	Comments
1.1	Does every display begin with a title or header that describes screen contents?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
1.2	Is there a consistent icon design scheme and stylistic treatment across the system?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
1.3	Is a single, selected icon clearly visible when surrounded by unselected icons?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
1.4	Do menu instructions, prompts, and error messages appear in the same place(s) on each menu?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
1.5	In multipage data entry screens, is each page labeled to show its relation to others?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
1.6	If overtype and insert mode are both available, is there a visible indication of which one the user is in?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
1.7	If pop-up windows are used to display error messages, do they allow the user to see the field in error?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
1.8	Is there some form of system feedback for every operator action?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	

Severity Scale:



- 0 - don't agree that this is a usability problem
- 1 - cosmetic problem
- 2 - minor usability problem
- 3 - major usability problem; important to fix
- 4 - usability catastrophe; imperative to fix



These could also be levels of severity

<https://wiki.library.oregonstate.edu/confluence/Heuristic+Evaluation+Checklist.pdf>

Heuristics

- Nielsen's heuristics have been modified to suit particular types of interfaces (e.g., web, mobile, etc)
- Nielsen recommends these to be supplements to the general ones
- There should be between 5-10 heuristics (less than 5 is too little, more than 10 may be too many for experts to remember)

Heuristics

- Nielsen's heuristics have been modified to suit particular types of interfaces (e.g., web, mobile, etc)
- Nielsen recommends these to be supplements to the general ones
- There should be between 5-10 heuristics (less than 5 is too little, more than 10 may be too many for experts to remember)

Heuristics

- Sometimes company guidelines are turned into heuristic by turning guidelines into short statements or questions that follow each guideline: Example (Nielsen's guidelines for navigation on the web):
 - G: avoid orphan pages
 - H: Are there any orphan pages? Where do they go?
- G: Avoid long pages with excessive white space that forces scrolling
 - H: Are there any long pages? Do they have lots of white space?
- G: Provide navigation support, such as a strong site map
 - H: is there any guidance? Maps, navigation bar, menus etc to help users around the site
- G: Provide a consistent look and feel for navigation and info design
 - H: Are menus used, named, and positioned consistently?
- Etc.

Excerpt from a Heuristics Worksheet

Heuristic Evaluation - A System Checklist

1. Visibility of System Status

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

#	Review Checklist	Yes	No	N/A	Comments
1.1	Does every display begin with a title or header that describes screen contents?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
1.2	Is there a consistent icon design scheme and stylistic treatment across the system?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
1.3	Is a single, selected icon clearly visible when surrounded by unselected icons?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
1.4	Do menu instructions, prompts, and error messages appear in the same place(s) on each menu?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
1.5	In multipage data entry screens, is each page labeled to show its relation to others?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
1.6	If overtype and insert mode are both available, is there a visible indication of which one the user is in?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
1.7	If pop-up windows are used to display error messages, do they allow the user to see the field in error?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
1.8	Is there some form of system feedback for every operator action?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	

Severity Scale:



- 0 - don't agree that this is a usability problem
- 1 - cosmetic problem
- 2 - minor usability problem
- 3 - major usability problem; important to fix
- 4 - usability catastrophe; imperative to fix



These could also be levels of severity

<https://wiki.library.oregonstate.edu/confluence/Heuristic+Evaluation+Checklist.pdf>

How to do a Heuristic Evaluation

There are three stages:

1. Briefing session
2. Evaluation period
3. Debriefing Session

How to do a Heuristic Evaluation

1. Briefing session: experts are explained what to do (prepared script is often used so everyone receives the same info/treatment)

How to do a Heuristic Evaluation

2. Evaluation period: each expert spends 1-2 hours independently inspecting the interface using the heuristics as guidance
 - They take at least 2 passes through the interface
 - First time – they get a feel for the flow of the interface
 - Second time – focus on specific elements in the interface and context of the whole product
 - It is useful to provide the evaluators with some pre-determined tasks (but they can do their own as well) – especially with earlier prototypes
 - Evaluator can take notes, or have them speak out loud and someone records their thoughts

How to do a Heuristic Evaluation

3. Debriefing Session: all the evaluators come together to discuss their finding and prioritize their problems they found and suggest solutions
 - Again someone should take notes

From last class - How to Evaluate

Two ways to evaluate designs

- With real users (usability studies with participants)
- Without real users (experts, analysis, inspections, etc.)

Evaluations with Users	Evaluations without Users
Controlled experiment	Cognitive Walkthroughs
Field studies	Heuristics
Interviews	Key-stroke
Questionnaires/surveys	Simulation
Observation	



Bug Bash by Hans Bjordahl



Bug Bash by Hans Bjordahl



Copyright 2005 Hans Bjordahl

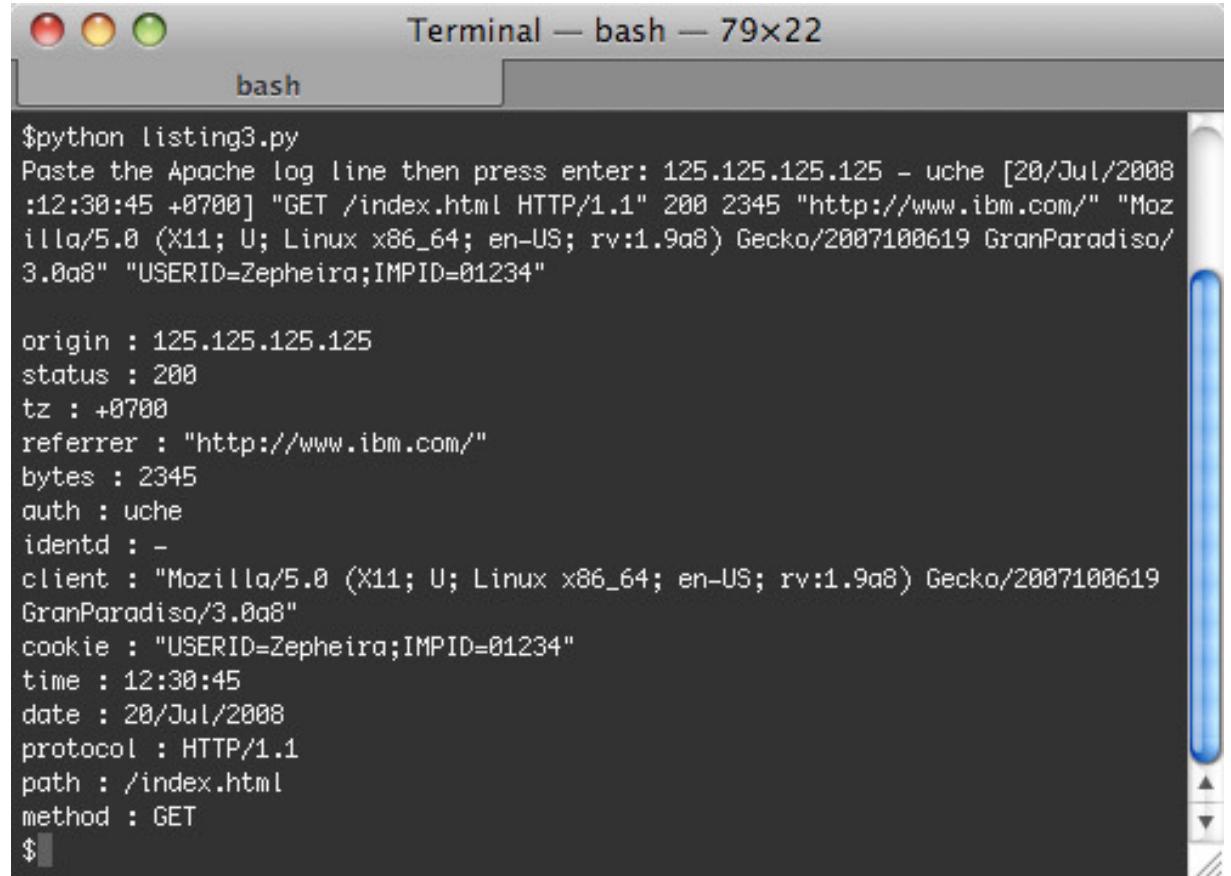


<http://www.bugbash.net/>

Using Logs to gather feedback

- Log actions
 - automatically collect user input on the computer
 - E.g., **keystrokes**, time to do actions, mouse clicks, scrolling
 - Capture user events on the Web
 - E.g., Server Logs and Client-side logs
 - URLs retrieved
 - time between interactivity (e.g., go to a new page)
 - how many times you revisit web pages
 - Links clicked on a page
 - Analyze search terms

Example: Server- side Log



The screenshot shows a Mac OS X terminal window titled "Terminal — bash — 79x22". The window has three colored window controls (red, yellow, green) at the top left. The title bar also displays "bash". The terminal window contains the following text:

```
$python listing3.py
Paste the Apache log line then press enter: 125.125.125.125 - uche [20/Jul/2008
:12:30:45 +0700] "GET /index.html HTTP/1.1" 200 2345 "http://www.ibm.com/" "Moz
illa/5.0 (X11; U; Linux x86_64; en-US; rv:1.9a8) Gecko/2007100619 GranParadiso/
3.0a8" "USERID=Zepheira;IMPID=01234"

origin : 125.125.125.125
status : 200
tz : +0700
referrer : "http://www.ibm.com/"
bytes : 2345
auth : uche
identd : -
client : "Mozilla/5.0 (X11; U; Linux x86_64; en-US; rv:1.9a8) Gecko/2007100619
GranParadiso/3.0a8"
cookie : "USERID=Zepheira;IMPID=01234"
time : 12:30:45
date : 20/Jul/2008
protocol : HTTP/1.1
path : /index.html
method : GET
$
```

Example: Client-side Log

Computer Captured Data						
eventid	window	tab	date	time	name	url
32	0	5	Wed Jul 16 2008	13:16:02	Clicked Toolbar Back to	http://www.booking.com/city/it/florence.en.html
33	0	5	Wed Jul 16 2008	13:16:06	Followed Link	http://www.booking.com/hotel/it/grand-baglioni.html?label=gog235jc;sid=910694d967dd42b1f54f0f60802b51
34	0	4	Wed Jul 16 2008	13:16:15	Clicked Toolbar Back to	http://www.google.ca/search?hl=en&q=florence+information&btnG=Google+Search&meta=
35	0	4	Wed Jul 16 2008	13:16:17	Followed Link	http://www.italyheaven.co.uk/florence.html
36	0	3	Wed Jul 16 2008	13:16:22	Followed Link	javascript:ManageMyBookings();
37	0	2	Wed Jul 16 2008	13:16:24	Switched away from task	
38	0	1	Wed Jul 16 2008	13:16:42	Switched into task	
39	0	1	Wed Jul 16 2008	13:16:43	Closed Tab	1
40	0	5	Wed Jul 16 2008	13:16:48	Switched into task	
41	0	5	Wed Jul 16 2008	13:16:49	Followed Link	http://www.booking.com/hotel/it/hotellondraflorence.html?label=gog235jc;sid=910694d967dd42b1f54f0f608
42	0	5	Wed Jul 16 2008	13:17:01	Stopped Task	CHI Conference (task16)
43	0	5	Wed Jul 16 2008	13:17:01	Stopped Task	CHI Conference-Reason:
44	0	5	Wed Jul 16 2008	13:17:01	Stopped Task	CHI Conference-Other Applications Used: none
45	0	5	Wed Jul 16 2008	13:17:01	Stopped Task	CHI Conference-Is Stopped Task16-Finished? no
46	0	5	Wed Jul 16 2008	13:17:04	Closed Tab	5
47	0	4	Wed Jul 16 2008	13:17:05	Closed Tab	4
48	0	3	Wed Jul 16 2008	13:17:06	Closed Tab	3
49	0	2	Wed Jul 16 2008	13:17:11	(Ctrl-T) Opened New Tab	
50	0	2	Wed Jul 16 2008	13:17:11	Tab 6 Started at	about:blank
51	0	6	Wed Jul 16 2008	13:20:36	Started New Task	book flight to Toronto (task17)
52	0	6	Wed Jul 16 2008	13:20:41	Stopped Task	book flight to Toronto (task17)
53	0	6	Wed Jul 16 2008	13:20:41	Stopped Task	book flight to Toronto-Reason:
54	0	6	Wed Jul 16 2008	13:20:41	Stopped Task	book flight to Toronto-Other Applications Used: none
55	0	6	Wed Jul 16 2008	13:20:41	Stopped Task	book flight to Toronto-Is Stopped Task17-Finished? no
56	0	6	Wed Jul 16 2008	13:20:54	Closed Window	
57	0	6	Wed Jul 16 2008	13:20:54	Done Session	
58	0	0	Wed Jul 16 2008	13:21:22	Started Session	
59	0	0	Wed Jul 16 2008	13:21:22	Started Window at Specified URL	www.google.com/ig
60	0	0	Wed Jul 16 2008	13:21:32	(Ctrl-T) Opened New Tab	
61	0	0	Wed Jul 16 2008	13:21:32	Tab 1 Started at	about:blank
62	0	1	Wed Jul 16 2008	13:24:01	(From Toolbar) Opened Bookmark (Under) ***	
63	0	1	Wed Jul 16 2008	13:24:06	(Ctrl-T) Opened New Tab	
64	0	1	Wed Jul 16 2008	13:24:06	Tab 2 Started at	about:blank

User Inputted Data

More on Heuristics

- http://www.useit.com/papers/heuristic/heuristic_list.html
- <https://www.usability.gov/how-to-and-tools/methods/heuristic-evaluation.html>

How to Evaluate

Two ways to evaluate designs

- With real users (usability studies with participants)
- Without real users (experts, analysis, inspections, etc.)

Evaluations with Users	Evaluations without Users
Controlled experiment	Cognitive Walkthroughs
Field studies	Heuristics
Interviews	Key-stroke
Questionnaires/surveys	Simulation
Observation	



Bug Bash by Hans Bjordahl

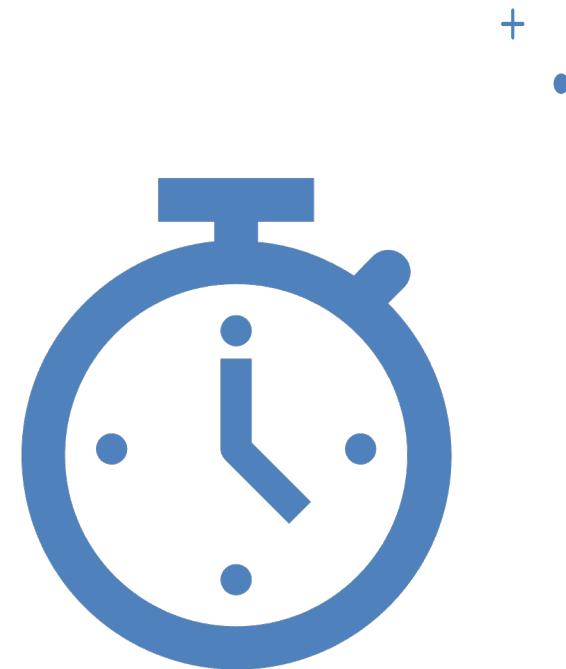


Controlled Study (aka Laboratory Study)

- Researcher tries to control as much as possible, such as
 - the physical environment (e.g., lighting, location)
 - The workspace
 - The equipment
 - The tasks
- The results tend to be more valid than other studies because the researcher has control over the “who, what, where, and how” of the experiment
- The users go through the same experience
 - Can potentially *isolate* what you are testing (e.g., font type, menu display, navigation methods, etc.)

Typical metrics

- Time to complete a task.
- Time to complete a task after a specified time away from the product (e.g, multi-session).
- Number and type of errors per task.
- Number of errors per unit of time.
- Number of navigations to online help or manuals.
- Number of users making a particular error.
- Number of users completing task successfully.



Other metrics?

- Not everything is performance driven...

User Preference

User Satisfaction

Likes/dislikes

Suggestions and opinions

Enjoyment

Learnability

Others?...

Metrics of interest

Usability

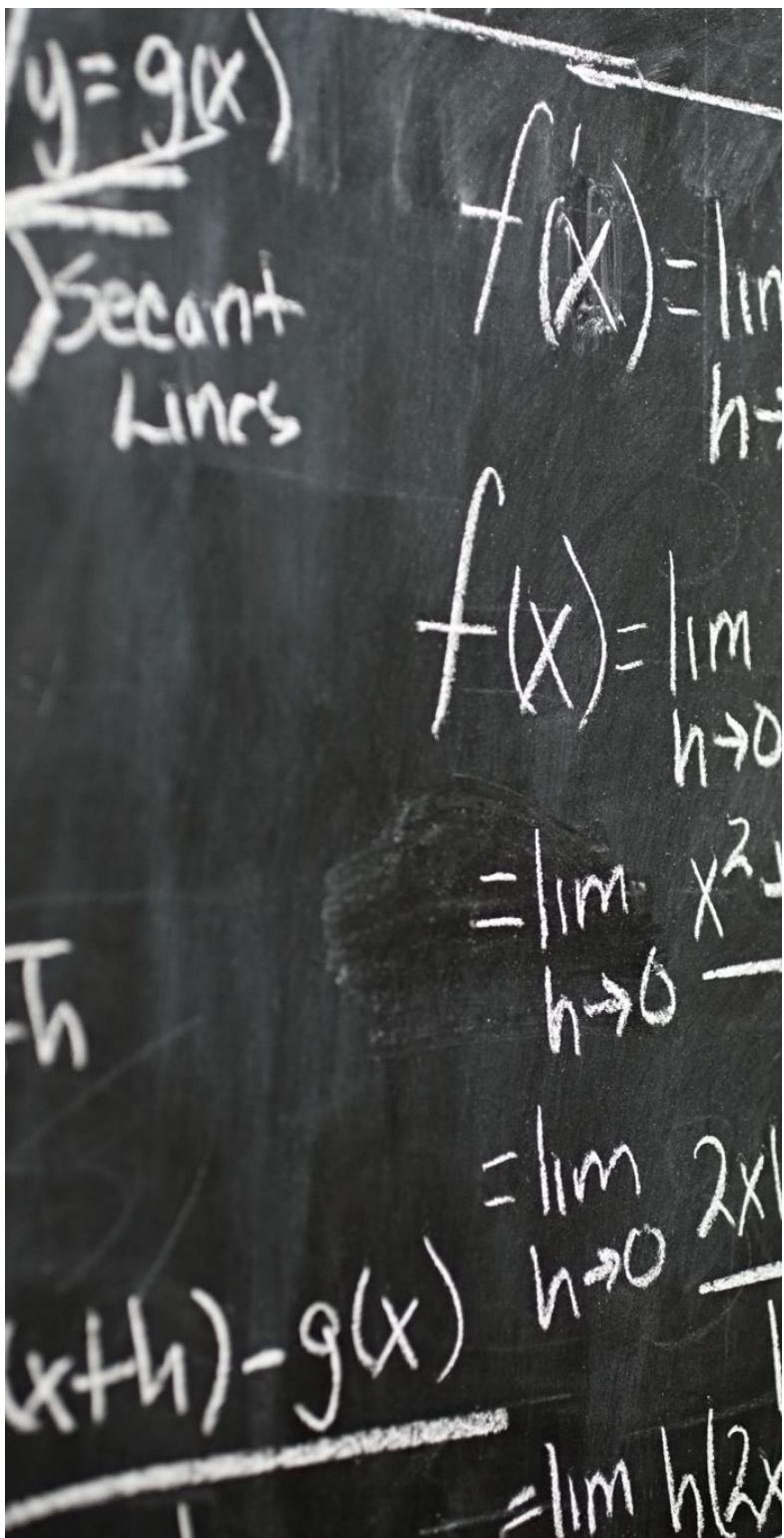
- Efficient
- Effective
- Safe
- Easy to learn
- Easy to remember
- Have good utility

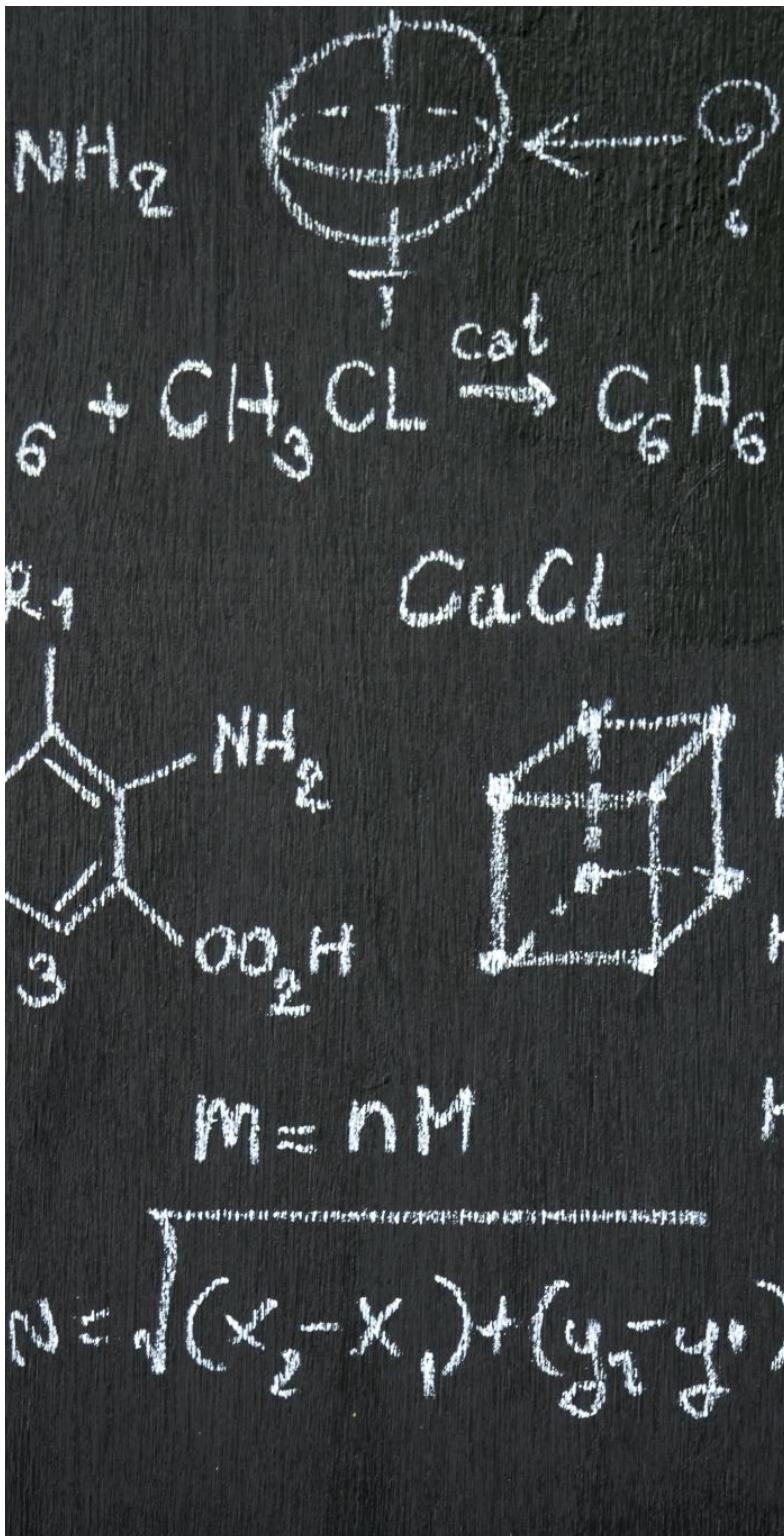
User Experience

- Support creativity
- Be motivating
- Be helpful
- Be satisfying to use

Controlled Studies

- Test for hypothesis – for example:
 H → reading text displayed in 12 pt Arial font is faster than reading text in 12 pt Times New Roman
- When you do a controlled study to evaluate this hypothesis (or others) you need to control for different factors – we will look at three





Controlled Studies: Factors to Control

1. Control for *testing variables* (this is what you are testing)
 - In this example, the researcher is testing the relationship between the two variables:
 - i. 12 pt Arial
 - ii. 12 pt Times New Roman

Testing Variables

- There are two types of variables that are involved in a controlled study:
 1. **Independent Variable/s** – this is what the researcher **manipulates** in the study
 - (e.g., font type – users read using Arial and with Times New Roman)
 2. **Dependent Variable/s** – this is dependent on the manipulated variable
 - (e.g., the time to read is a dependent variable; it is dependent on the font type)
 - If the assumptions are correct in the hypothesis, then the researcher should see a difference between reading based on the fonts used

Outside Variables

2. Control for Outside Variables and Environment

- A big challenge is to determine the other variables that may influence the results of the study (you want to ensure that everyone experiences the same thing)
- In this example, you would want to make sure that you keep certain conditions consistent between variables and participants:
 - Font colour
 - Contrast
 - Screen resolution, etc...
 - WHY??

To make results valid!

e.g., if participants use two different computer screens, each with a different resolution, and you see a difference between reading times with fonts you will not know if it's due to the font types or the resolution

Order

3. Control Order

- You should *randomize* the order that your participants use your independent variables (e.g., using the font types (Arial and Times New Roman in this example) among the participants)
- In most cases, ‘randomized’ is a controlled random order
- For example: half your participants would use Arial then Times New Roman, and half would use Times then Arial
- (if you have large enough numbers you can randomize the order but normally you need to ensure that each order combination is considered and incorporated)

Order

- Control Example:
 - If you were testing 3 different font types:
Font 1 (F1) Font 2 (F2) and Font 3 (F3)

WHY DO THIS??

Participant	Order		
	1 st	2 nd	3 rd
1	F1	F2	F3
2	F2	F3	F1
3	F3	F1	F2

To avoid issues with learning biases

- This ensures that each participant views each font type in a different order and it ensures that the font type used first/second/third was different

How to organize users/participants

- If you are comparing different interfaces (conditions) you can organize your users to test these different and multiple conditions by either doing:
- Between-subjects study design: different people test each condition, so that each person is only exposed to a single user interface.
- Within-subjects (or repeated-measures) study design: the same person tests all the conditions (i.e., all the user interfaces).

Between Study vs Within Study

Between-subjects study design:

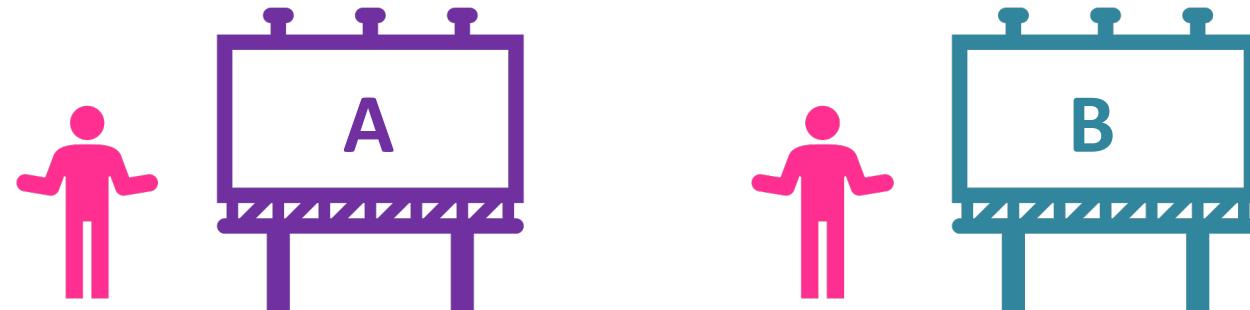
- This means that you assign different users to only test one condition (e.g., they are only exposed to one interface design)

Within-subjects (aka repeated-measures) study design:

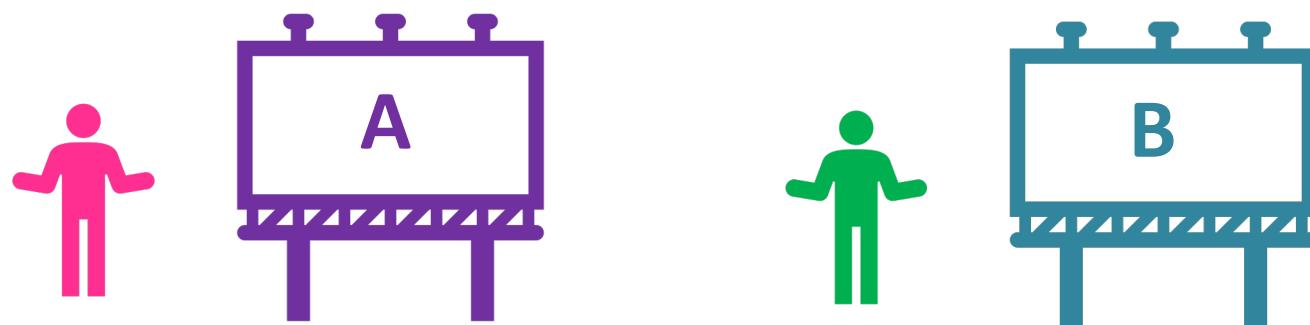
- This means that the same user tests all the conditions (i.e., they try out all the interface options).

Hypothesis: Users will be faster finding photos on a large screen display using Interface A compared to Interface B.

Within Study
Design (same user
tries both)



Between Study
Design (half users
try A and the other
B)



Advantages of Between Study

- Minimizes the learning effect (only use interface)
- Less time to do study (only use one interface)
- Can be easier to set up if don't need to randomize all the variables

Advantages of Within Study

- Because users experience all interfaces, it might be easier to determine differences
- Can use less users because they test all variables (e.g., interfaces)

For Complex Evaluations

- May use a combination of both
- For example, if there are a lot of variables to consider:

2 interface design
options

5 different task
types

2 different device types
(e.g., tabletop vs wall)

A study design approach:

Between subjects for devices (e.g., Group 1 uses table top and Group 2 uses wall)

Within subjects:

Group 1 tries all 5 tasks using both interfaces (will need to control for order)

Group 2 tries all 5 tasks using both interfaces (will need to control for order)

User Tasks



When you are performing the laboratory study, you need to figure out what you will have the users do (tasks) which will be done by all users



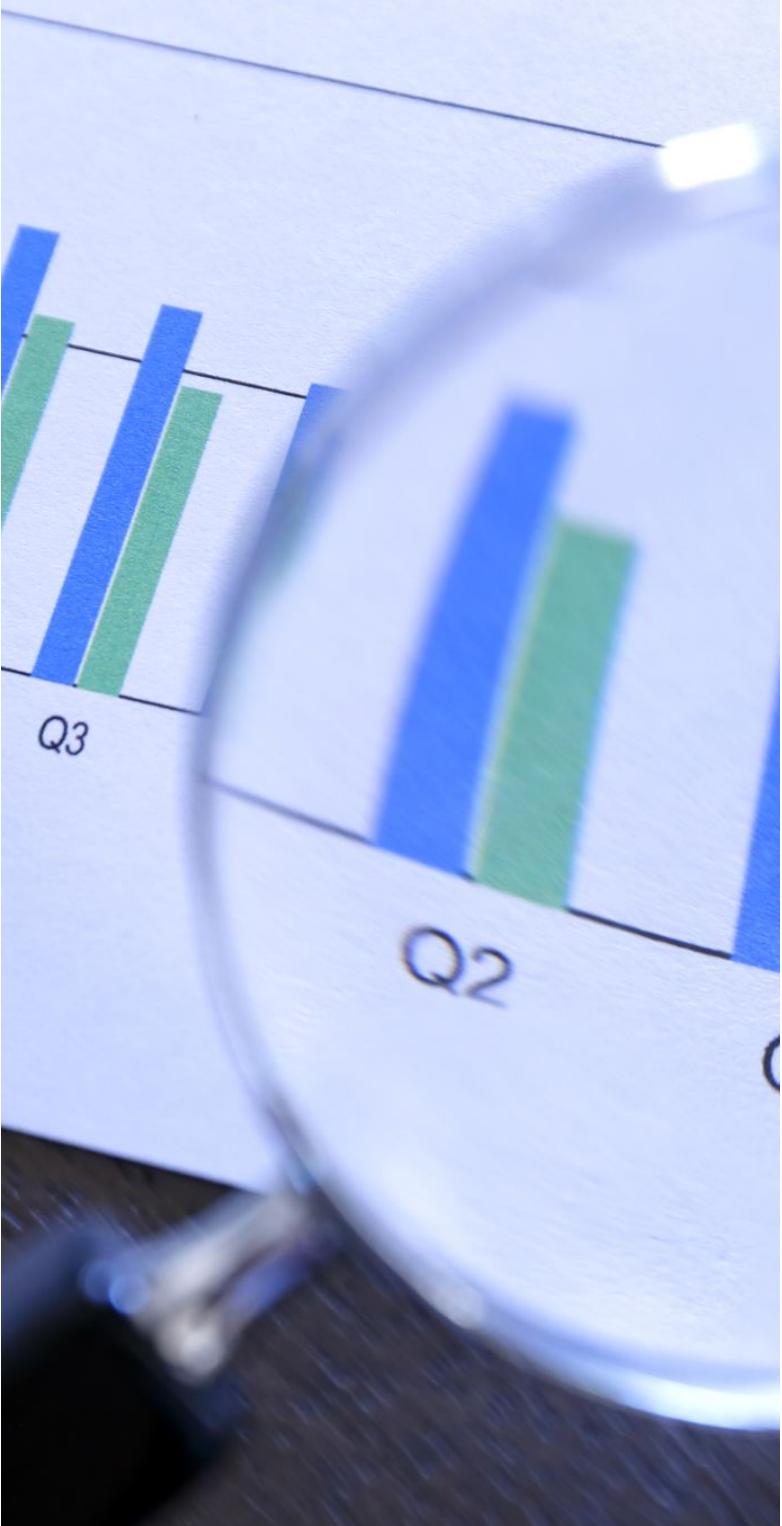
Your tasks should help you answer your research question/s or prove your hypothesis



Tasks should be clear and have a meaningful goal



You should pilot (do a pre-study test with small number of users – a *dress rehearsal*) your study before the real one in order to identify confusing or too difficult tasks or to find bugs in prototypes/systems



In your study you will collect data

Two types of data collected in studies:

- Quantitative Data
 - numerical data
 - Examples:
 - Logged data, times, counts of interactions
- Qualitative data
 - observations, opinions, reasoning behind doing something or results, feelings
 - Examples:
 - User comments, open-ended questions, researcher's notes

Metrics of interest

Usability

- Efficient
- Effective
- Safe
- Easy to learn
- Easy to remember
- Have good utility

User Experience

- Support creativity
- Be motivating
- Be helpful
- Be satisfying to use

Example

- Hypothesis:
 - It will be **faster** to search on Google using a desktop than an iPhone.
- How to measure this??

Example

- Hypothesis:
 - Users **will prefer** searching Google using a desktop than an iPhone.
- How to capture this?

Example

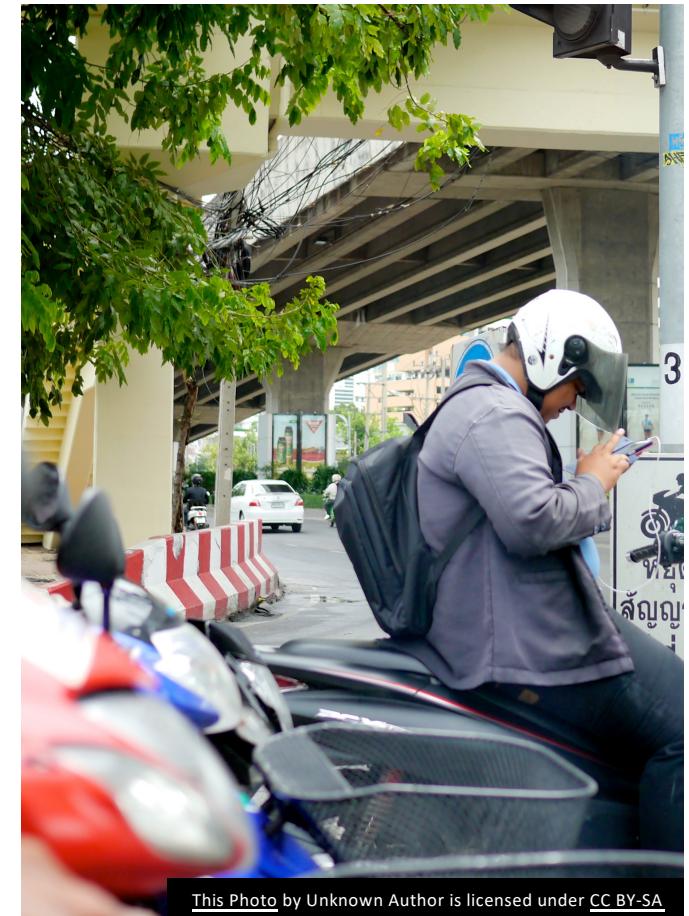
- Hypothesis:
 - Users will be **more effective** searching Google using a desktop than an iPhone.
- How to capture this?

Controlled/Laboratory Study

- Advantages:
 - Researcher can control the environment and independent variables (e.g., font type) which makes it easier to test specific factors (e.g., a hypothesis) and collect data
- Disadvantages
 - The decrease in context (e.g., using an interface artificially) within a real environment and user motivation (e.g., made up tasks) makes it more difficult to generalize the results (e.g., you can't be positive that it will adapted or used in the way you plan once the user is in the real world)

Field Studies/Field Evaluations

- With Field Studies the environment is not controlled, and tasks are not set (as with controlled studies) – rather users are encouraged to do their day-to-day activities as they normally would
 - e.g., if they are working on a task on the web then we would expect there to be interruptions such as using the phone, task-switching like reading emails, etc.
- Can be done with direct observation or indirect observation



This Photo by Unknown Author is licensed under CC BY-SA

Direct Observation

- Watch in person how someone does something
- Provides context, lots of detail, based on what “actually happens” not what is expected to happen
- People may or may not know (e.g., asked to watch people in a business before designing software vs. watching strangers at the ATM machines)
- Can bias the situation (act differently being watched - eg The Hawthorne effect)
- Takes time, expensive to do many (\$\$ and time)
- Can realistically only do one pass of watching (due to expense)
- Based on notes and personal observation, so may miss something

Indirect Observation

- Used somehow to record the information (not there in person) e.g., video recording, audio, logging software
- Provides a permanent record and review later
- Can watch or capture actions/tasks over a longer period (e.g., several days/weeks)
- Users may act more naturally because distance between user and observer but still a bit obtrusive
- Issues of privacy/security with having a permanent record
- Need lots of time to go through the data (e.g., to watch the videos, go through the logs)

Field study vs field evaluation

- In the true form a field study refers to a purely observational study with no real involvement in a person's environment (we watch to learn how people do things – can be in person, video, or logs)
- But in Computer Science we tend to do more *Field Evaluations*:
 - adds something new into the environment (e.g., a new tool or interface) and then we observe usually with participant feedback
 - Often in our field, we try to understand how a product or prototype is adapted and used by people in their everyday lives
 - But you need to be careful to ensure the safety of your participants

Field Studies

- Unlike controlled studies where the results are easy to capture and measure e.g., time and satisfaction (and thus compare between participants), Field Studies/Evaluations results can be messy (due to real life situations which we have no control over)
 - e.g., time doesn't work as a measure because in a FS task participants may not do the task continuously from start to finish (e.g., phone calls, answer emails, etc)

Data Collected

Review: Two types of data collected in studies:

- Quantitative Data
 - numerical data (Logged data, times, counts of interactions, etc.)
- Qualitative data
 - observations, opinions, reasoning behind doing something or results, feelings (User comments, suggestions, a researcher's notes, etc.)

Data Collected

- In a FS you capture a lot of qualitative data (e.g., observations, opinions, reasoning behind doing something or results, feelings)
 - You can capture some quantitative data (numerical data) e.g., logging interactions - count times a person goes on line to visit help, revisits to web pages
- In a laboratory/controlled study you tend to capture more quantitative data (numerical data – e.g., times to do a task or a rating on a questionnaire)
 - You can also get some qualitative data as well with post-study questionnaires, interviews...

Field Studies

- Advantages:
 - You can evaluate how people think about something, interact with a product and integrate a product into their own setting
 - Therefore, you can get a better sense of how successful products will be used in the real world
- Disadvantage:
 - But it is more difficult to test a specific hypothesis or assertion about a product or interface, or test particular details of an interface (e.g., how fast you can do something) because we can not control the where, when, and how (double-edged sword)

Interviews (review)

- You ask users (usually face to face) about their experience and opinions using software or about their tasks to gain information for new development
- Can be Structured
 - ask the same questions in the same order (rigid)
 - Thorough and consistent but can not probe answers
- Can be Semi-structured
 - use an interview guide with questions to ask can ask questions in a different order and probe answers
- Can be Open or Flexible
 - No structure
 - Can be difficult to run but you may get more information as it is user driven (need experienced interviewer)
 - Harder to compare between users

Interviews

- Interviews:
 - Advantages:
 - In depth information and personal experience
 - If record (video or audio) have a permanent record
 - Disadvantages:
 - Timely and costly (therefore limited number)
 - Need to ensure don't ask leading questions or you could bias the results
 - If record the conversation, takes time to transcribe

Focus Groups

Focus Groups

- Interviews tend to be one to one (interviewer to interviewee)
- But you can also interview people in groups (e.g., marketing, and social sciences)
- The method assumes that individuals develop opinions within a social context by talking with others
- Usually 3-10 people involved and is led by a facilitator (often trained to do focus groups)

<https://www.youtube.com/watch?v=ZLw0YXcseG0>



<http://www.youtube.com/watch?v=OKbMQIv-4aE&feature=related>

<http://www.youtube.com/watch?v=pcj7QT0Abk8&feature=related>

Focus Groups

Advantages:

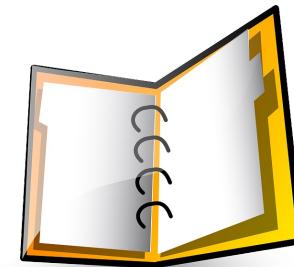
- If you create a supportive environment where people feel comfortable, you will get invaluable feedback and opinions
- You can use the dynamics of a group to promote discussion and sometimes can get richer information than with one on one interviews

Disadvantages:

- The success of the focus group largely lies with the facilitator:
 - needs to direct the flow of conversation or topics (otherwise the group could stuck on one topic only)
 - needs to encourage quiet people to participate and stop verbose people from dominating the group discussion
- If you record the session, you will have a permanent record but you will also need time to translate and analyze the data

Diary Studies

- Helps with longitudinal studies (e.g., with field studies/field evaluations using observations)
 - Evaluation must be situated in real user activities, in the wild
 - But cannot be with a participant 24/7
- Ask participants to self-report data capture using techniques
 - Paper and pencil notes
 - Online reporting, such as in a blog
 - Cellphone voice-mail messages/ pictures/video
 - Pocket digital voice recorder
- User takes notes to document within long-term usage:
 - Problems
 - Experiences
 - Occurrences that indicate meaningfulness
- Need to decide when recording happen
 - Is it event driven (decided by user or system notification?)
 - Is random (prompted by system)
 - Is regular intervals (e.g., every night before bed)
 - Need to consider the user and their daily activities (e.g., can they do this while at work, is the information sensitive so don't want them to record in public)



Example of Closed Diary



UX2 Diary Study 2010 - (online form)

Welcome to UX2 diary study. We are currently reviewing digital services provided through University of Edinburgh's Library and would like to understand how you use these services to achieve your goals.

* Required

Today's diary

Please step through the following sections giving as much detail as you can. You do not need to fill in all the fields at the same time so if you prefer to keep the form open in the background as you work, please be assured that your session will not time out. The form will retain all your information until you are ready to submit your diary entry.

This is a diary entry for: *
(enter your name here)

Your Location? *
(where were you doing your research)

Start Time? *
(note the time that you started)

Duration? *
(how long did you spend on your research)

Activities

Tell us WHY you were searching

What were you searching for today? *
(provide a brief description on the purpose of your task)

On a scale of 1 to 9, how important was this activity to you?

1 2 3 4 5 6 7 8 9

<http://tinyurl.com/2b4n7bs> 84

Forms of analysis

- Track temporal patterns
- Look for semantic patterns in visual data
- Combine with interview analysis

Diary Studies

- Advantages:
 - Since can't be with people 24-7, allows for recording of actual activities, emotions, and events
 - Reporting can be immediate while details are fresh
- Disadvantages
 - Not as objective as direct observation but still a practical solution
 - Users have to be willing and committed to recording these events in a manner that has been asked (e.g., to take a photo, once a day, etc.) for the entire period.
 - May need measures to remind users (but how often?)

Questionnaires/Surveys

- Can be used for post-study sessions to get opinions and feedback from participants
- Can be a stand-alone study
 - Don't need to perform in person - easier to reach large numbers of people
 - Need a lot of people to respond to generalize results
- Ensure have pertinent questions and well-structured questions that answer a specific question/s
 - Questions shouldn't be ambiguous and need to make sense (there isn't someone there to 'clarify' answers)
 - Questions must help you answer your research question or test your hypothesis
 - Need to be careful that you don't ask misleading questions



Stone, D., Jarrett, C., Woodroffe, M., Minocha, S. (2005). User Interface Design and Evaluation: Morgan Kaufmann, California.

Sharp, Rogers, Preece. (2007). Interaction Design: beyond human-computer interaction (2nd edition): John Wiley & Sons, England.



Types of Questions

- While questionnaires tend to be a mix of both, often interviews and focus groups are more open-ended questions (though still some mixture)
- Some questions make more sense in a questionnaire/survey
- Open-ended questions
 - What do you think about....
 - Tell me about a time when
 - Walk me through how you do
- Closed-ended questions
 - Yes/No
 - Multiple choice
 - Likert Scale



Styles of Questions: Open-ended

- asks for opinions
- good for general subjective information
 - but difficult to analyze rigorously

for example,

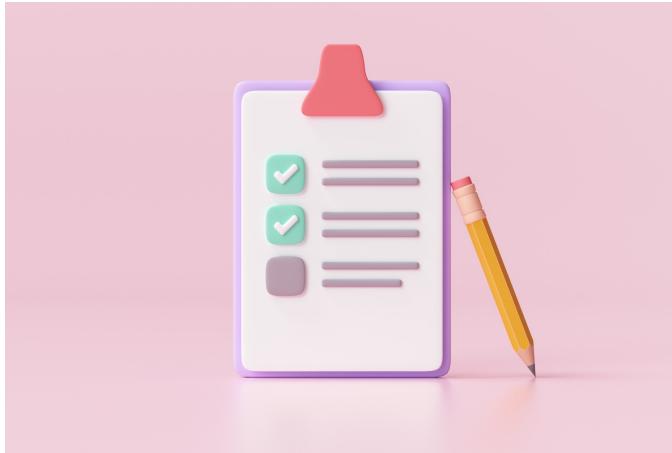
- “Can you suggest any improvements to the interface?”

Open-ended Questions



- These questions do not restrict the user to certain choices and leaves the answer up to the user
- E.g., description of user experiences, opinions, feedback or suggestions
- Examples:
 - What do you use your mobile phone for?
 - What suggestions do you have to make the application easier to use?
 - What do you like/dislike about the website?
 - How would you improve the website?

Styles of Questions: Closed



- restricts responses by supplying the choices for answers
- can be easily analyzed ...
- but can still be hard to interpret, if questions / responses not well designed!
 - alternative answers should be very specific

Do you use computers at work:

often sometimes rarely

In your typical workday, do you use computers:

- over 4 hrs a day
- between 2 and 4 hrs daily
- between 1 and 2 hrs daily
- less than 1 hr a day

Close-ended Questions



Checkboxes and Ranges

offer users a choice of several answers or a range of answers

- e.g., education level, employment type

- e.g., some people don't like to enter their exact age or income (so often see these a ranges)

Age:

- 18-25
- 26-35
- 36-45
- 46+

Income:

- < \$25,000
- \$26,000 - \$35,000
- \$36,000- \$45,000
- \$46,000 +

Select all that you do on the Web:

- Read/send e-mail
- Use IM
- Go on Facebook
- Read the news
- Shop

Styles of Questions: Scalar - Likert Scale

- measure opinions, attitudes, and beliefs
- ask user to judge a specific statement on a numeric scale
- scale usually corresponds to agreement or disagreement with a statement
 - purpose to get a range of responses to a question that can be compared across participants
 - E.g., preferences, likes, usability
 - most common type of scale: Likert which measures attitudes and beliefs
 - This makes Likert scales good to evaluate user satisfaction with products/interfaces etc.

Characters on the computer screen are hard to read:

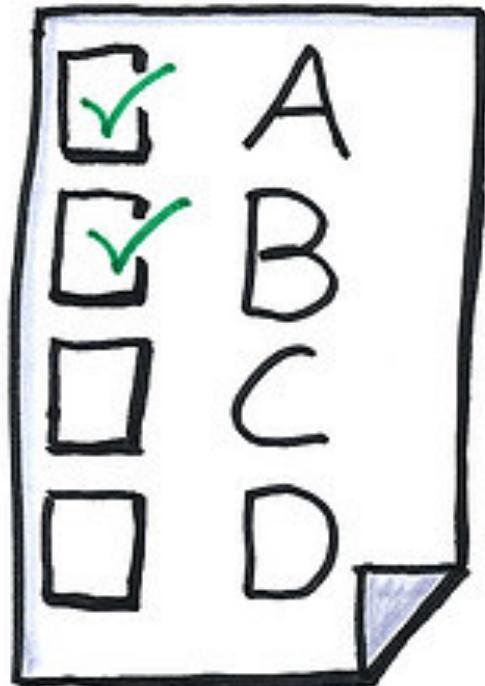
strongly agree

1 2 3 4 5

strongly disagree

Styles of Questions: Multi-Choice

- respondent offered a choice of explicit responses that they can answer with *just one choice or can check multiple choices.*



How do you most often get help with the system? (tick one)

- on-line manual
- paper manual
- ask a colleague

Which types of software have you used?
(tick all that apply)

- word processor
- data base
- spreadsheet
- compiler

Styles of Questions: Ranked



- respondent places an ordering on items in a list
- useful to indicate a user's preferences
- forced choice

Rank the usefulness of these methods of issuing a command

(1 most useful, 2 next most useful..., 0 if not used)

 2 command line

 1 menu selection

 3 control key accelerator

Ranked Questions

- One problem with ranked questions is how close is the 1st place with 2nd place ranking?

Rank the usefulness of these methods of issuing a command

(1 most useful, 2 next most useful..., 0 if not used)

Easier to analyze

__2__ command line

__1__ menu selection

__3__ control key

1. menu

3. Control key

1

2. Command line

3

0

More accurate
assessment

Closed questions: advantages & disadvantages

Advantages:

- easy and quick to answer
- the answers of different respondents are easier to code, analyze, and compare
- the response choices can clarify question meaning for respondents
- respondents are more likely to answer about sensitive topics

Disadvantages:

- They can suggest ideas that the respondent would not otherwise have
- Respondents with no opinion or no knowledge may answer anyway
- It is confusing if many response choices are offered
- Misinterpretation of a question can go unnoticed
- Clerical mistakes or marking the wrong response is possible
- They force respondents to give simplistic responses to complex issues or make choice they would not make in the real world

Open questions: Advantages & Disadvantages

Advantages

- they permit an unlimited number of possible answers.
- respondents can answer in detail and can qualify and clarify responses
- unanticipated findings can be discovered
- they permit adequate answers to complex issues
- they permit creativity, self-expression, and richness of detail
- they reveal a respondent's logic, thinking process, and frame of reference

Disadvantages

- different respondents give different degrees of detail in answers
- responses may be irrelevant or buried in useless detail
- comparisons and statistical analysis become difficult
- coding responses is difficult· articulate and highly literate respondents have an advantage
- a greater amount of respondent time, thought, and effort is necessary
- respondents can be intimidated by questions
- answers take up a lot of space in the questionnaire and/or take more time to answer

Class Activity – Using Heuristics

- Go to this site: <http://blowoff.com>

Explore at least three parts of the site:

- Start by evaluating the home page;
- Then evaluate the page for one product, such as BlowOff Duster;
- Finally, try to buy the product and evaluate the checkout page.
- Record the usability problems you find. Justify every observation by naming one or more usability heuristics (design principles) that it violates. Assign a severity rating to each problem (cosmetic, minor, major, or catastrophic).
- Try and include at least one positive usability comment for each heuristic as well, again justifying it by naming one or more heuristics.

- <http://ics.colorado.edu/techpubs/pdf/92-17.pdf>
- <http://www.allbusiness.com/media-telecommunications/internet-www/10629523-1.html>
- <http://www.cc.gatech.edu/classes/cs3302/documents/cog.walk.html>
- <http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-87/html/#ref-01>
- <http://hcibib.org/tcuid/chap-4.html#e3>
- http://digitalcommons.utep.edu/cgi/viewcontent.cgi?article=1009&context=cs_papers
- Sharp, H., Rogers, Y., Preece, J. (2007). Interaction Design: beyond human-computer interaction.(2nd Edition). England: John Wiley & Sons, Ltd.
- <http://www.bugbash.net/comic/7.html>
- <http://widemile.wordpress.com/category/why-test/>
- <http://www.userinsight.com/solutions/usability-testing/>