City University Interaction Lab

Listen
Compare
Observe
Measure

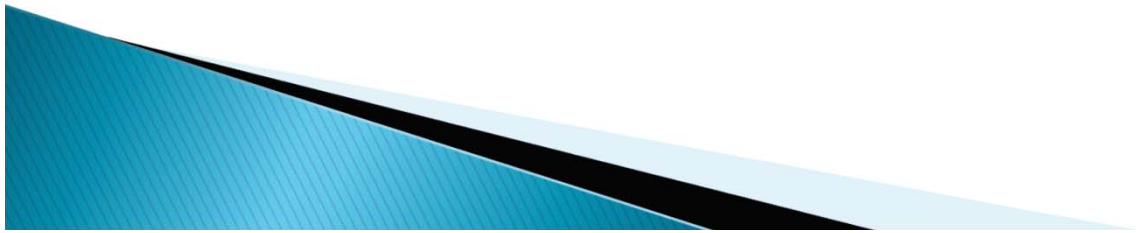
Unit 6. Evaluating Interface Designs

Interfaces Persona Computador

Depto. Sistemas Informáticos y Computación. UPV

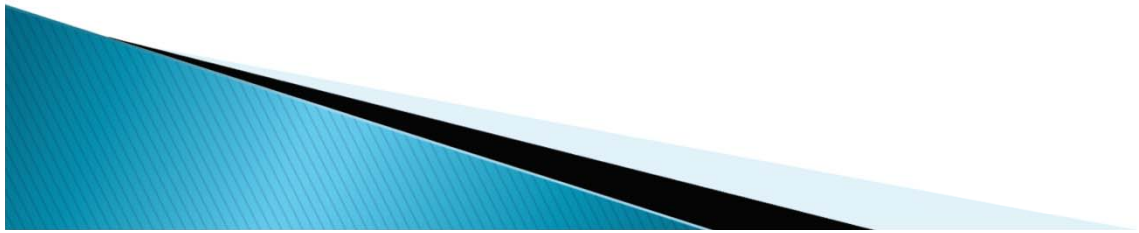
Unit Goals

- Learn the different types of usability evaluation techniques
- Understand the characteristics of each type of evaluation, and be able to select the proper technique for each project
- Know how to organize a user study session
- Know the different types of expert inspection techniques
- Understand the types of evaluation performed along the life cycle of a system



Outline

- Introduction
- Types of Evaluation
- User Studies
- Expert Reviews
- Other studies
- Acceptance Tests
- Evaluation During Active Use



Introduction

- Evaluating our interfaces is essential
  - Helps to understand the user experience with the system and, where there are difficulties, to find ways of improving it.
- Depending on the project, an appropriate amount of resources should be assigned to testing (5-20%)
- Usability tests have to be carried out during the whole development process, not only at the end
  - e.g., early in the design process, several paper mockups can be shown to users for testing and discussion
  - and take into account their suggestions!

Introduction

- In critical systems, both high load situations and even partial failure situations should be tested
  - and, sometimes, this cannot be done in the laboratory, and has to be done in its real context
- There are multiple evaluation methodologies
  - depending on the application type, type of interface, type of users…
  - how many users will test the system? which type of test? evaluation experts or real users? how can all the tasks of an application be tested?...

Introduction

- Remember that usability is not an abstract concept. It can be measured and evaluated
- The evaluation should measure how the final product adheres to the usability requirements
  - Qualitative Usability Requirements
    - Desired features. They can be subjective and sometimes hard to measure. “The users should be able to order an item easily and without assistance”
  - Quantitative Usability Requirements/Usability Metrics
    - The requirement is expressed with a number “Any page of the site should load in less then 10 s in a 3G connection”, “4 out of 5 novice users must rate the product as ‘easy to use’ or ‘very easy to use’ on a 5-point scale”

Introduction

- Levels for evaluating a usability metric
  - Current
  - Best case
  - Planned
  - Worst case
- For example:
  - The users need 4 minutes to complete a task with the existing system (current level). Ideally, we would like to halve it (best case). Probably 3 minutes is acceptable (planned). Anything longer than 3 minutes and 50 seconds is unacceptable (worst case)

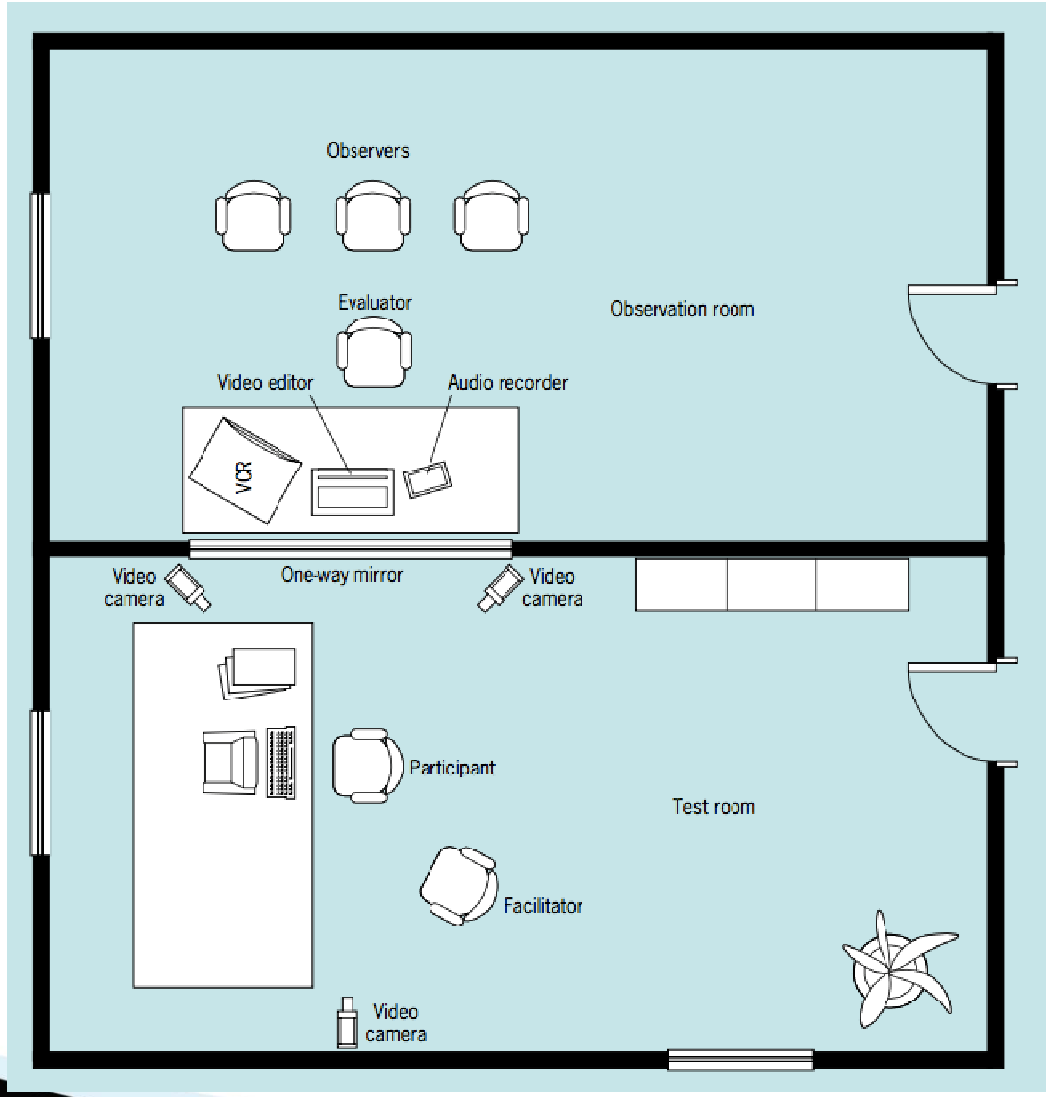
Types of evaluation

- What:
  - Find as many usability problems as possible: diagnostics evaluation
  - Assess the extent to which a system meets its requirements: measurement evaluation
- When:
  - During the development: formative evaluation
  - At the end: summative evaluation
- How:
  - Early, with low fidelity prototypes, informal: exploratory evaluation
  - At the end, verify system meets requirements, formal experiment: validation evaluation
  - Choosing one between several options, statistical analysis: comparison evaluation
- Who:
  - Users: User observations
  - Experts: Expert inspection
  - Other

User Studies

- Choose your users
  - How many? Usually 5 users is enough
  - Who: ideally a real user, a representative of a user profile, a usability expert or a domain expert
- Depending on where the study is performed:
  - Users' own environment: field studies
  - Other: controlled studies

Usability laboratories



Usability laboratories

- Typical usability laboratory: subject's space

Unidad de Investigación Acceso de la Universitat de València

Usability laboratories

- Typical usability laboratory: observers' space

http://www.hq.nasa.gov/pao/portal/usability

Usability laboratories

- Field tests and portable labs
  - The usability lab is brought to the place where the final system will be used
  - It is important to capture the largest amount of information in each execution
  - Other option is to release beta versions

to a high number of users, and ask for comments

noldus.com

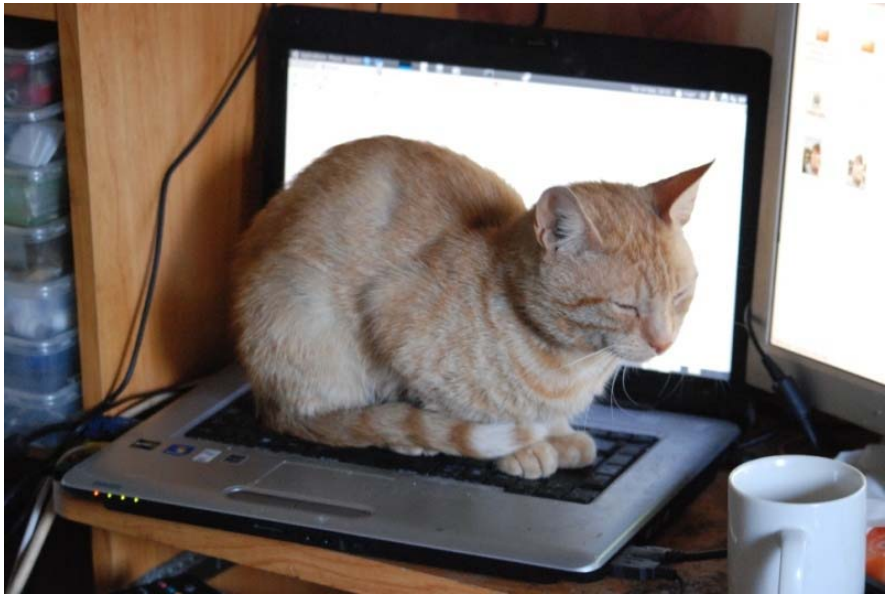
Types of usability tests

- Remote usability testing
  - For web-based applications, a large number of users can test the system from the place where they will use it
  - Participants can be recruited by e-mail from client databases or on-line forums
  - Tests can be synchronous or asynchronous
  - Advantages: large number of participants, inexpensive, tests the users hardware
  - Disadvantages: less control on the user behavior and it is difficult to capture their reactions
  - Some studies have found that this type of tests find more problems than traditional techniques

Types of usability tests

- Can-you-break-this tests
  - Videogame developers pioneered this type of tests, where the user is asked to “break” the system
  - These stress tests help to make the applications more robust

flickr.com/photos/quazie



flickr.com/photos/susan_g

Types of usability tests

- Competitive usability testing
  - Compare a new interface to previous version, or with the interface of a competitor
  - Compare the time it takes to complete a task, or the error rate in both systems

Types of usability tests

- Competitive usability testing
  - Metrovalencia: http://www.metrovalencia.es/wordpress/?page_id=298
  - Results of survey. May 2012
  - *Question 4: If you had to buy an AB area round ticket, which of the following two sequences of screens (A and B) would be better? Why?*

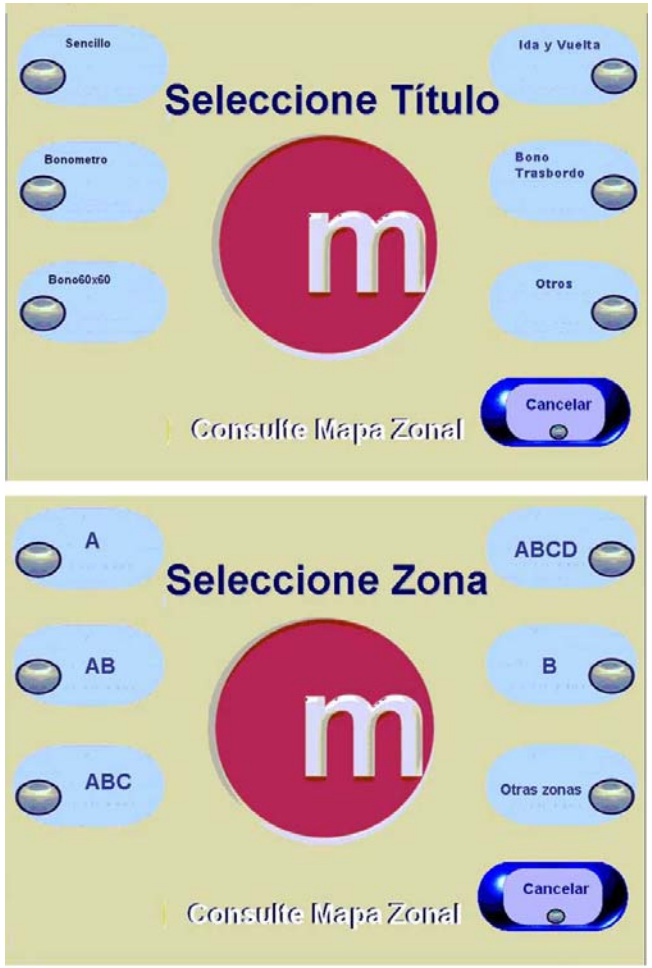
Option A **Option B**

Types of usability tests

- Competitive usability testing

Option A

Option B



Preparing the experiment

- Before the experiment, a detailed plan has to be agreed upon, including:
  - what to measure,
  - number, types and sources of the participants,
  - duration (30-90 min), tasks to be carried out by the subjects, and
  - the content of the questionnaires and interviews
- Run a pilot test with a small number of subjects (1-3)
  - Tests that all details of the evaluation are taken care of
  - Perform the test in the same location and as similarly as possible to the actual tests
  - Analyze and interpret the data, to ensure that all the required information is collected

Structure of a Session

- Welcome
- Sign a consent form and maybe an Non-disclosure agreement
- Recruitment screener
- Use task scenarios (adapted to users' language) for specifying the instructions for the users
  - Select the most important ones
- Post-session discussion
  - Review the recordings asking about their thoughts and questions
  - Questionnaires
- Incentive

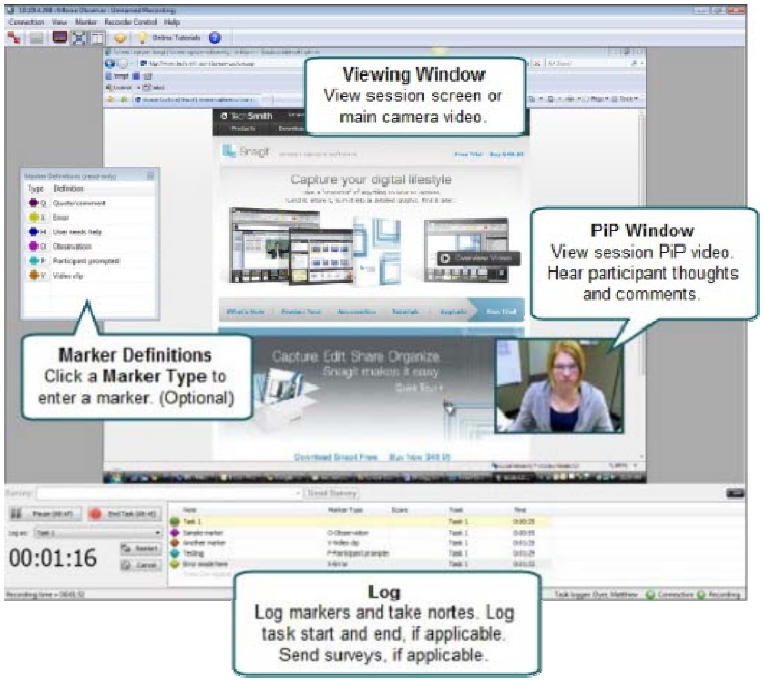
Structure of a Session

- Welcome
  - Participants should always be treated with respect, and let them know that **it is not them** who are being evaluated
  - They should also be informed about:
    - the purpose of the study,
    - what they will be doing and for how long. How to ask for a break,
    - who will review the recording and their use after the study,
    - a statement of confidentiality and how the anonymity of the participant is preserved,
    - the risks of taking the test,
    - the fact that participation is voluntary and she can withdraw at any time with no penalty
    - a way to contact with questions
- Recruitment Screener: ensure that the participant fits to our expected profile
  - backgrounds in computing, experience with the task, motivation, education level, ability with the natural language
  - eyesight, dominant hand, age, gender

During the Test

Measuring time data

- Use a stopwatch
- Record a time stamp with every comment written during the session
  - Or use a key logger, that records every keystroke or mouse click. Problem: relate written comments with each event
  - Or specialized software for usability evaluations (Ovo Solo, Noldus The Observer XT, TechSmith Morae…)



techsmith.com

During the Test

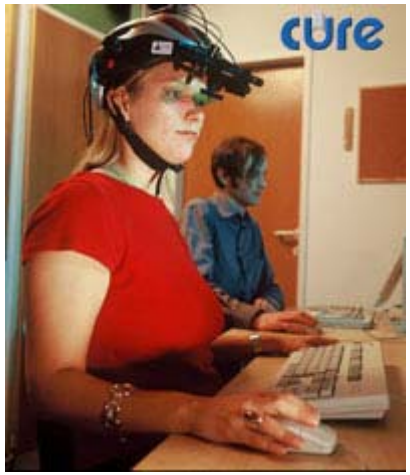
- A usability testing technique consists of asking the participant to think aloud
  - The observer should facilitate the communication and prompt for what the user is thinking. The observer must not help the subject
  - http://www.youtube.com/watch?v=QckIzHC99Xc
  - Advantages
    - Immediate feedback on the participant's opinions about the interface and any problems or surprises.
    - It can help users to focus and concentrate during the evaluation session.
    - Useful for collecting qualitative data
  - Disadvantages:
    - Some participants can find thinking aloud unnatural and distracting.
    - Thinking aloud can slow the participant's thought processes, artificially reducing their performance and error rates
    - It can become very exhausting for the user

During the Test

- Retrospective protocol: ask users for comments about their actions after the test
  - Advantages
    - Does not interfere with performance measurements
    - Useful for collecting quantitative data
  - Disadvantages
    - The participant may forget the reason she did some action
    - Some participants can be intimidated by the cameras

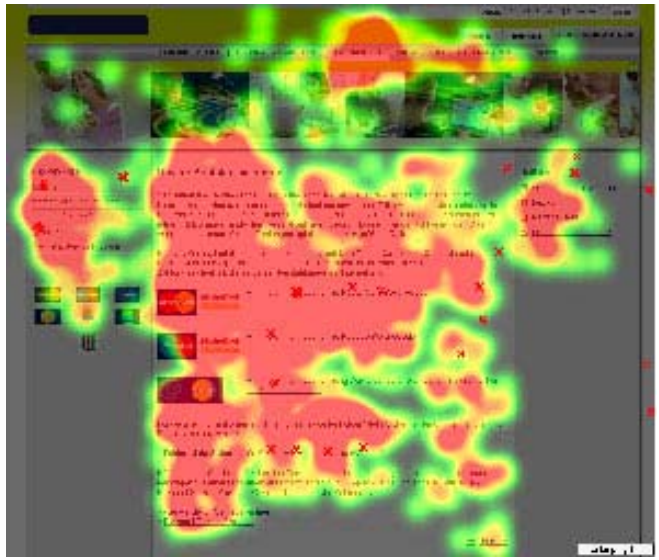
Recordings

- Recording the participants is valuable for reviewing later their reactions, errors, how they work, etc.
- Eye tracking system compute the areas of the screen that receive more attention



cure.at

tobii.com



cure.at

Recordings

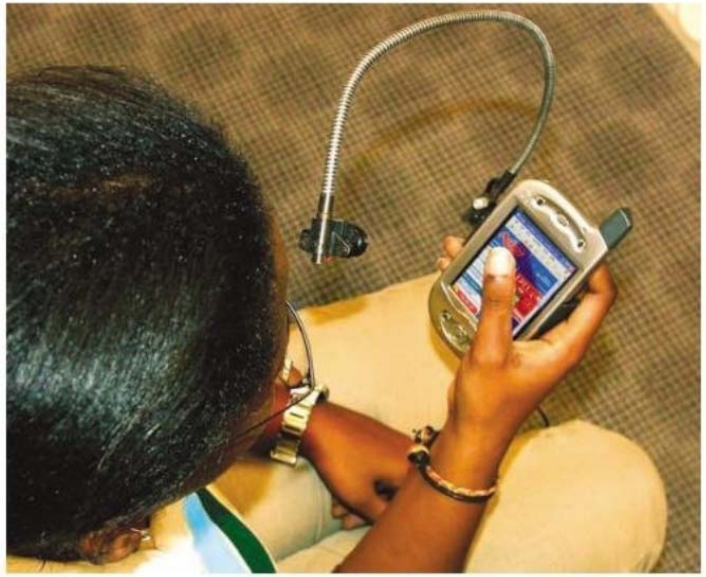


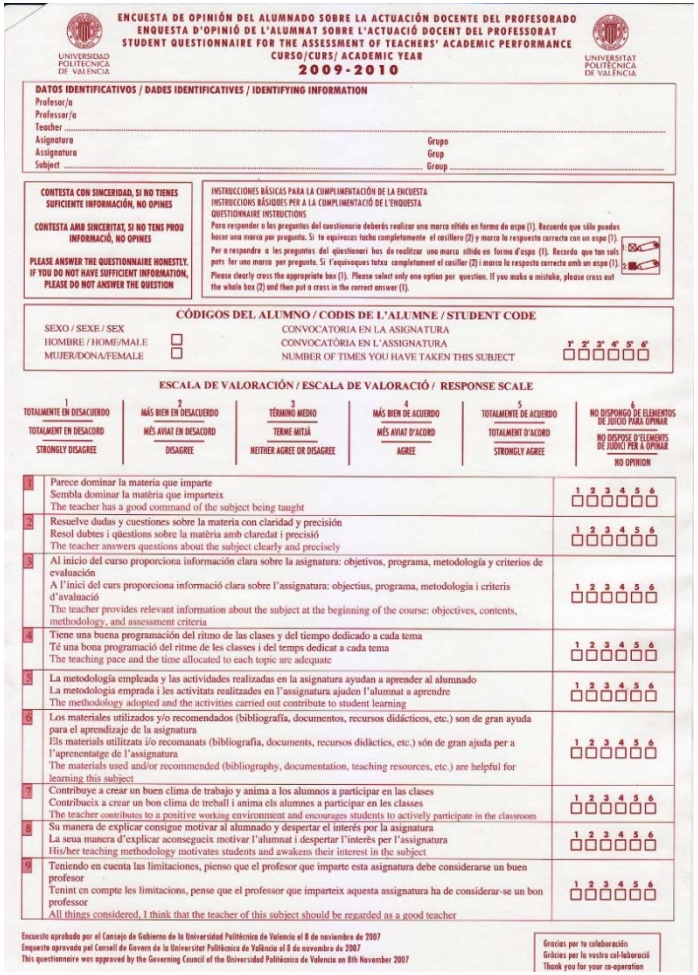
FIGURE 4.5
Special mobile camera to track and record activities on the screen of a hand-held device (http://www.tracksys.co.uk/product-details.php?id=9).

FIGURE 4.4
Portable lab with eye tracking (http://www.mangold-international.com/en.html).

Source: Designing the User Interface: Pearson New International Edition: Strategies for Effective Human-Computer Interaction. *Shneiderman, Ben* | *Cohen, Maxine; Plaisant, Catherine* ISBN 1-292-03701-6

Questionnaires

- Familiar, inexpensive method to capture the user's or expert's opinion
- It allows to poll thousands of users
- The keys to success are having clear goals and developing focused items
- Paper or on-line



ENCUESTA DE OPINIÓN DEL ALUMNADO SOBRE LA ACTUACIÓN DOCENTE DEL PROFESORADO
ENQUESTA D'OPINIÓ DE L'ALUMNAT SOBRE L'ACTUACIÓ DOCENT DEL PROFESSORAT
STUDENT QUESTIONNAIRE FOR THE ASSESSMENT OF TEACHERS' ACADEMIC PERFORMANCE
CURSO/CURS/ ACADEMIC YEAR
2009-2010

UNIVERSIDAD POLITÉCNICA DE VALENCIA
UNIVERSITAT POLITÈCNICA DE VALÈNCIA

DATOS IDENTIFICATIVOS / DADES IDENTIFICATIVES / IDENTIFYING INFORMATION
Profesor/a
Professor/a
Teacher
Asignatura / Assignatura / Subject
Grupo / Grup / Group

CONTESTA CON SINCERIDAD, SI NO TIENES SUFICIENTE INFORMACIÓN, NO OPINES
CONTESTA AMB SINCERITAT, SI NO TENS PROU INFORMACIÓ, NO OPINES
PLEASE ANSWER THE QUESTIONNAIRE HONESTLY. IF YOU DO NOT HAVE SUFFICIENT INFORMATION, PLEASE DO NOT ANSWER THE QUESTION

INSTRUCCIONES BÁSICAS PARA LA CUMPLIMENTACIÓN DE LA ENCUESTA
INSTRUCCIONS BÀSIQUES PER A LA CUMPLIMENTACIÓ DE L'ENQUESTA
QUESTIONNAIRE INSTRUCTIONS
Para responder a las preguntas del cuestionario deberás realizar una marca nítida en forma de aspa (1). Recuerda que sólo puedes hacer una marca por pregunta. Si te equivocas tacha completamente el casillero (2) y marca la respuesta correcta con un aspa (1).
Per a respondre a les preguntes del qüestionari has de realitzar una marca nítida en forma d'aspa (1). Recorda que tan sols pots fer una marca per pregunta. Si t'equivoques ratlla completament el casiller (2) i marca la resposta correcta amb un aspa (1).
Please clearly cross the appropriate box (1). Please select only one option per question. If you make a mistake, please cross out the whole box (2) and then put a cross in the correct answer (1).

CÓDIGOS DEL ALUMNO / CODIS DE L'ALUMNE / STUDENT CODE
SEXO / SEXE / SEX
HOMBRE / HOME/MALE
MUJER/DONA/FEMALE
CONVOCATORIA EN LA ASIGNATURA
CONVOCATÒRIA EN L'ASSIGNATURA
NUMBER OF TIMES YOU HAVE TAKEN THIS SUBJECT
1ª 2ª 3ª 4ª 5ª 6ª

ESCALA DE VALORACIÓN / ESCALA DE VALORACIÓ / RESPONSE SCALE

1	2	3	4	5	6
TOTALMENTE EN DESACUERDO / TOTALMENT EN DESACORD / STRONGLY DISAGREE	MÁS BIEN EN DESACUERDO / MÉS AVIAT EN DESACORD / DISAGREE	TÉRMINO MEDIO / TERME MITJÀ / NEITHER AGREE OR DISAGREE	MÁS BIEN DE ACUERDO / MÉS AVIAT D'ACORD / AGREE	TOTALMENTE DE ACUERDO / TOTALMENT D'ACORD / STRONGLY AGREE	NO DISPONGO DE ELEMENTOS DE JUICIO PARA OPINAR / NO DISPOSE D'ELEMENTS DE JUDICI PER A OPINAR / NO OPINION

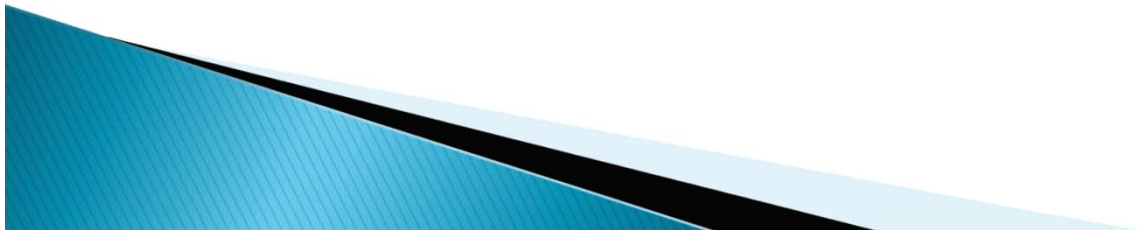
1	Parece dominar la materia que imparte Sembla dominar la matèria que imparteix The teacher has a good command of the subject being taught	1 2 3 4 5 6
2	Resuelve dudas y cuestiones sobre la materia con claridad y precisión Resol dubtes i qüestions sobre la matèria amb claredat i precisió The teacher answers questions about the subject clearly and precisely	1 2 3 4 5 6
3	Al inicio del curso proporciona información clara sobre la asignatura: objetivos, programa, metodología y criterios de evaluación A l'inici del curs proporciona informació clara sobre l'assignatura: objectius, programa, metodologia i criteris d'avaluació The teacher provides relevant information about the subject at the beginning of the course: objectives, contents, methodology, and assessment criteria	1 2 3 4 5 6
4	Tiene una buena programación del ritmo de las clases y del tiempo dedicado a cada tema Té una bona programació del ritme de les classes i del temps dedicat a cada tema The teaching pace and the time allocated to each topic are adequate	1 2 3 4 5 6
5	La metodología empleada y las actividades realizadas en la asignatura ayudan a aprender al alumnado La metodologia emprada i les activitats realitzades en l'assignatura ajuden l'alumnat a aprendre The methodology adopted and the activities carried out contribute to student learning	1 2 3 4 5 6
6	Los materiales utilizados y/o recomendados (bibliografía, documentos, recursos didácticos, etc.) son de gran ayuda para el aprendizaje de la asignatura Els materials utilitzats i/o recomanats (bibliografia, documents, recursos didàctics, etc.) són de gran ajuda per a l'aprenentatge de l'assignatura The materials used and/or recommended (bibliography, documentation, teaching resources, etc.) are helpful for learning this subject	1 2 3 4 5 6
7	Contribuye a crear un buen clima de trabajo y anima a los alumnos a participar en las clases Contribueix a crear un bon clima de treball i anima els alumnes a participar en les classes The teacher contributes to a positive working environment and encourages students to actively participate in the classroom	1 2 3 4 5 6
8	Su manera de explicar consigue motivar al alumnado y despertar el interés por la asignatura La seua manera d'explicar aconsegueix motivar l'alumnat i despertar l'interès per l'assignatura His/her teaching methodology motivates students and awakens their interest in the subject	1 2 3 4 5 6
9	Teniendo en cuenta las limitaciones, pienso que el profesor que imparte esta asignatura debe considerarse un buen profesor Tenint en compte les limitacions, pense que el professor que imparteix aquesta assignatura ha de considerar-se un bon professor All things considered, I think that the teacher of this subject should be regarded as a good teacher	1 2 3 4 5 6

Encuesta aprobada por el Consejo de Gobierno de la Universidad Politécnica de Valencia el 8 de noviembre de 2007
Enquesta aprovada pel Consell de Govern de la Universitat Politècnica de València el 8 de novembre de 2007
This questionnaire was approved by the Governing Council of the Universidad Politécnica de Valencia on 8th November 2007

Gracias por tu colaboración
Gràcies per la vostra col·laboració
Thank you for your co-operation

Questionnaires

- Advantages
  - Harder to forget to ask something
  - Comparing answers from different participants is easier, since all of them see the same questions
  - Quantitative data, such as “3 of the 4 participants said the interface was easy to navigate” can be collected
  - Progress is demonstrated by improved scores on subsequent surveys
- Disadvantages
  - It is difficult to design a good questionnaire
  - Closed questions are easier to analyze, but do not give the reason why the user selected it



Questionnaires

Types of questions

- Likert scale. The user has to assign an agreement level to a statement : strongly agree, agree, neutral, disagree, strongly disagree
  - Examples:
    - Improves my performance in book searching and buying
    - Enables me to search and buy books faster
    - Makes it easier to search for and purchase books
    - Increases my productivity in searching and purchasing books
  - It typically uses a 5-level scale, but 7 or 9 are also possible (always symmetrical)

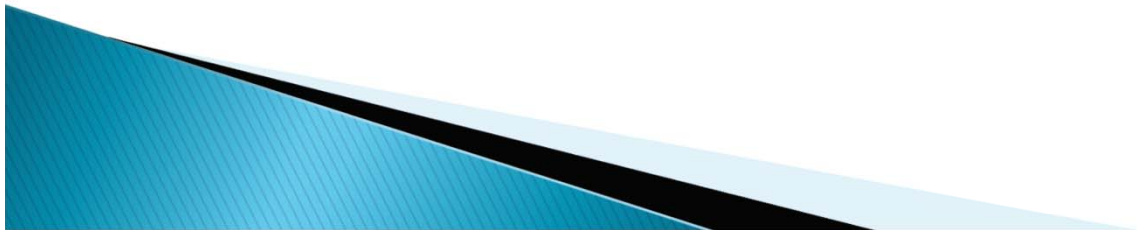
Questionnaires

Types of questions

- Bipolar items to describe users' reactions to using a system:
  - pleasing vs. irritating
  - simple vs. complicated
  - concise vs. redundant…
- Other options, rating of reactions:
  - Hostile 1 2 3 4 5 6 7 Friendly
  - Vague 1 2 3 4 5 6 7 Specific
  - Misleading 1 2 3 4 5 6 7 Beneficial
  - Discouraging 1 2 3 4 5 6 7 Encouraging

Questionnaires

- There are many predesigned, validated questionnaires that can be used:
  - QUIS Questionnaire for User Interaction Satisfaction
  - SUS System Usability Scale
  - CSUQ Computer System Usability Questionnaire
  - WAMMI Website Analysis and MeasureMent Inventory
  - …
  - More info about questionnaires in http://garyperlman.com/quest/



Questionnaires

- QUIS: *Questionnaire for User Interaction Satisfaction* (http://lap.umd.edu/quis)
  - Designed to evaluate user subjective satisfaction with respect to specific aspects of the interface
  - Studies nine factors of the interface:
    - display factors, terminology and system feedback, learning factors, system features, technical manuals, on-line tutorials, multimedia, teleconference and software installation
  - Nine-point scale
  - The questionnaire has to be adapted to the features of each interface

Questionnaires

- QUIS: Example
  - PART 1: System Experience
    - How long have you worked on this system?

❑ less than 1 hour
❑ 1 hour to less than 1 day
❑ 1 day to less than 1 week
❑ 1 week to less than 1 month
❑ 1 month to less than 6 months
❑ 6 months to less than 1 year
❑ 1 year to less than 2 years
❑ 2 years to less than 3 years
❑ 3 years or more

  - PART 6: Learning

6.1 Learning to operate the system

difficult easy
1 2 3 4 5 6 7 8 9 NA

6.1.1 Getting started

difficult easy
1 2 3 4 5 6 7 8 9 NA

6.1.2 Learning advanced features

difficult easy
1 2 3 4 5 6 7 8 9 NA

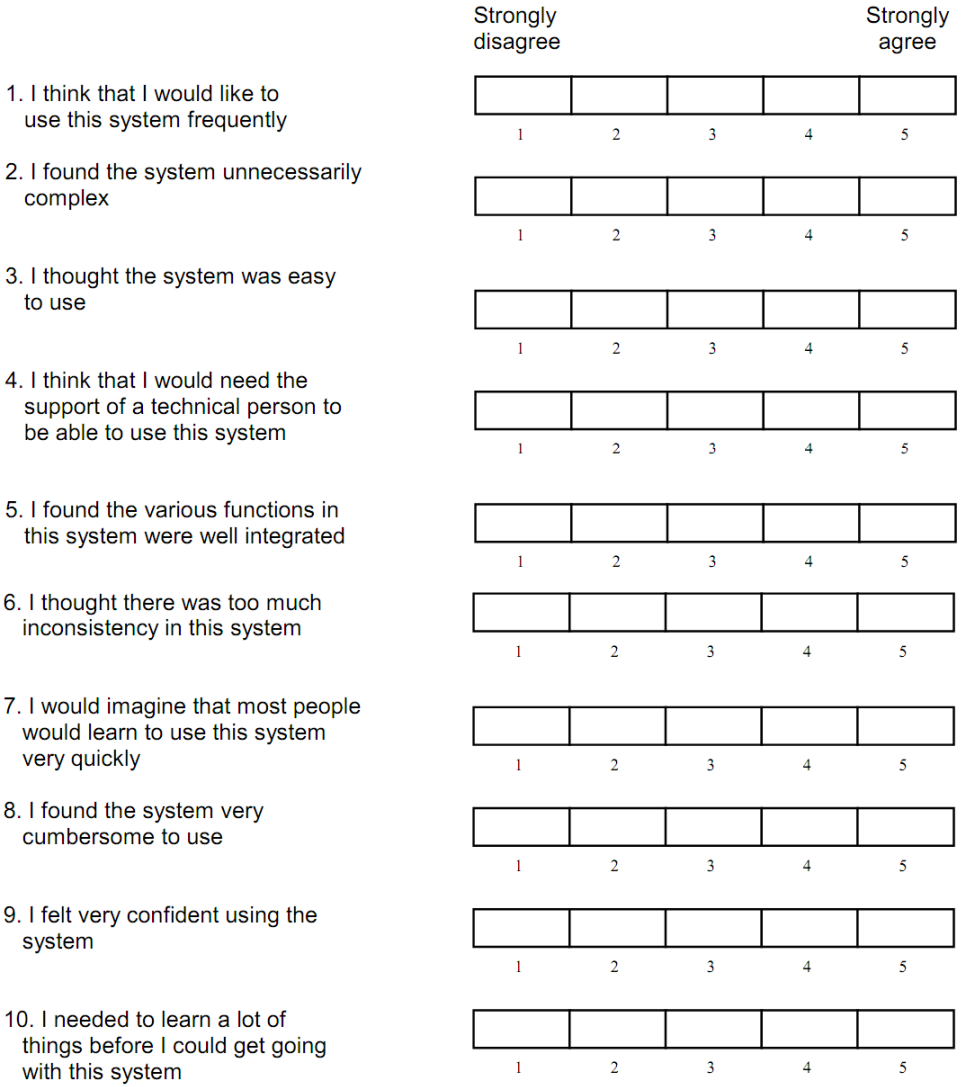
6.1.3 Time to learn to use the system

slow fast
1 2 3 4 5 6 7 8 9 NA

Questionnaires

- *System Usability Scale* (SUS)
  - Simpler than QUIS
  - 10 statements to which participants respond using a 5-point scale
  - Half of the questions are positively worded and the other half negatively worded

System Usability Scale



Brooke, John. SUS: A quick and dirty usability scale. Usability Evaluation in Industry, 1996

Questionnaires

- *Computer System Usability Questionnaire* (CSUQ)
  - 19 statements to which participants respond using a 7-point scale

1. Overall, I am satisfied with how easy it is to use this system.

STRONGLY AGREE	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**STRONGLY DISAGREE**

COMMENTS:

2. It is simple to use this system.

STRONGLY AGREE	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**STRONGLY DISAGREE**

COMMENTS:

3. I can effectively complete my work using this system.

STRONGLY AGREE	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**STRONGLY DISAGREE**

COMMENTS:

4. I am able to complete my work quickly using this system.

STRONGLY AGREE	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**STRONGLY DISAGREE**

COMMENTS:

5. I am able to efficiently complete my work using this system.

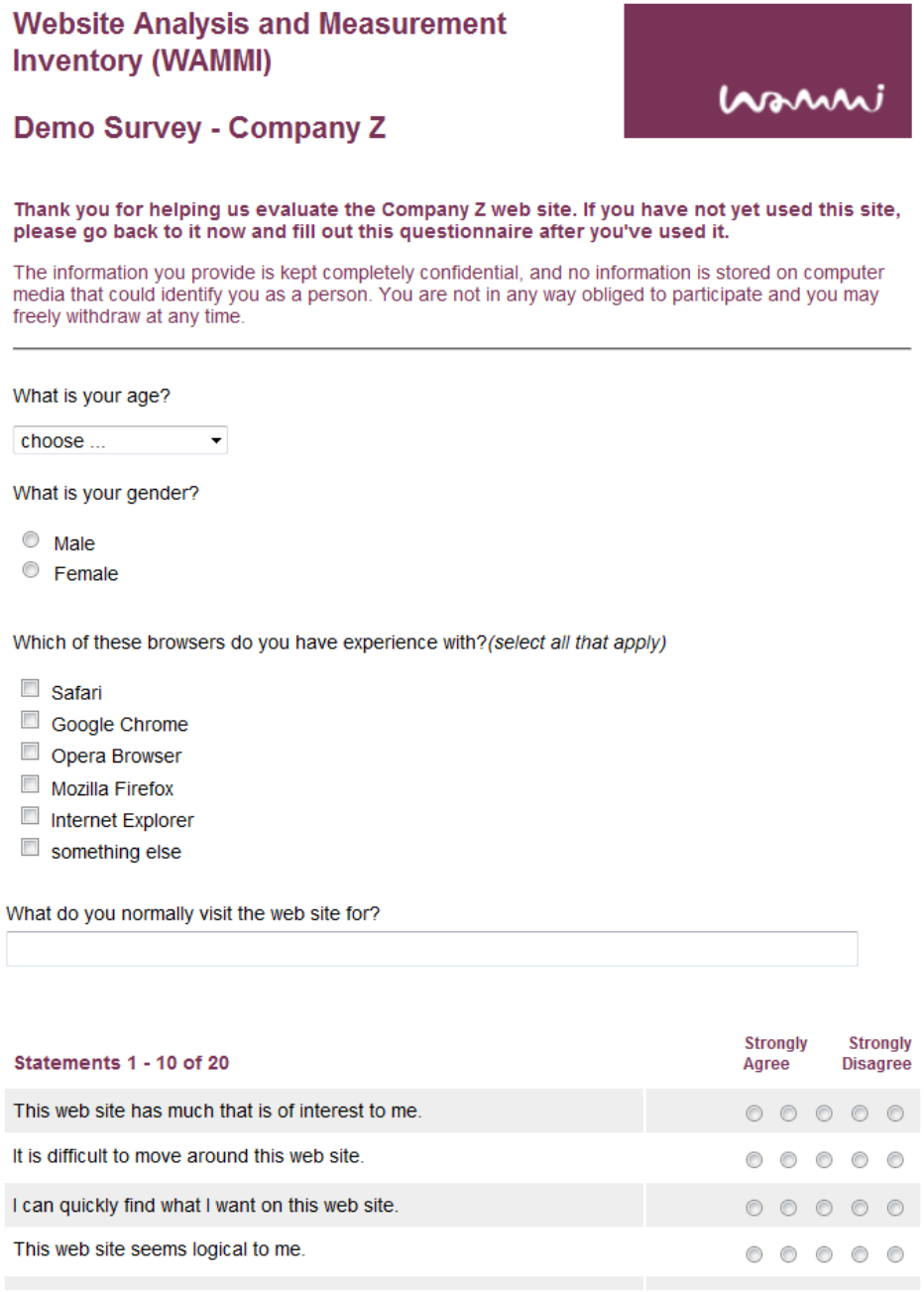
STRONGLY AGREE	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**STRONGLY DISAGREE**

COMMENTS:

J. R. Lewis. IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. Technical Report 54.786

Questionnaires

- *Website Analysis and MeasureMent Inventory* (WAMMI)
  - Web-based evaluation service
  - 20 statements to which participants respond using a 5-point scale
  - It provides web support to run the survey, and then it generates the final report



Website Analysis and Measurement Inventory (WAMMI)

Demo Survey - Company Z

Thank you for helping us evaluate the Company Z web site. If you have not yet used this site, please go back to it now and fill out this questionnaire after you've used it.

The information you provide is kept completely confidential, and no information is stored on computer media that could identify you as a person. You are not in any way obliged to participate and you may freely withdraw at any time.

What is your age?

choose ...

What is your gender?

- Male
- Female

Which of these browsers do you have experience with?*(select all that apply)*

- Safari
- Google Chrome
- Opera Browser
- Mozilla Firefox
- Internet Explorer
- something else

What do you normally visit the web site for?

Statements 1 - 10 of 20	Strongly Agree				Strongly Disagree
This web site has much that is of interest to me.					
It is difficult to move around this web site.					
I can quickly find what I want on this web site.					
This web site seems logical to me.					

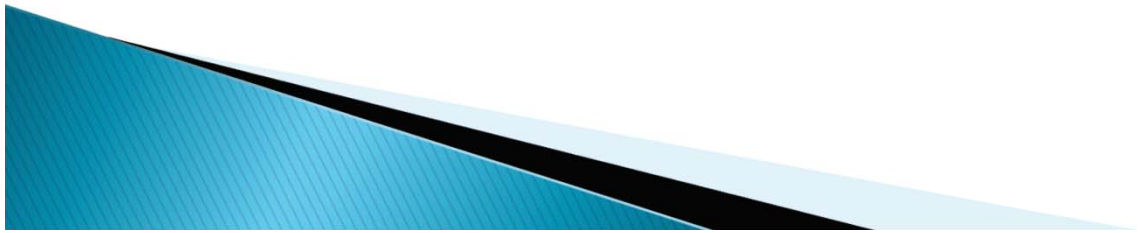
http://www.wammi.com/samples/index.html

Analysis of the Results

- After the session, you will have collected:
  - Background data about participants, notes by the evaluators, audio or video recordings, data collection forms, quantitative data on times, errors and other usability metrics, quantitative (and qualitative) data from pre- and post-session questionnaires, retrospective protocols, list of usability problems found…
- A usability defect is a usability problem in the user interface that can lead to confusion, error, delay or failure to complete some task

Summarizing Quantitative Data

- Tabulations, charts, and rankings
- Descriptive statistics (mean, median, mode)
  - Be careful: suppose a user needed 5 seconds to complete a task, another user required 6 seconds, and a third user 55 seconds
    - Mean: 22 seconds
- Inferential statistics (tests of statistical significance)
  - Requires a high number of participants



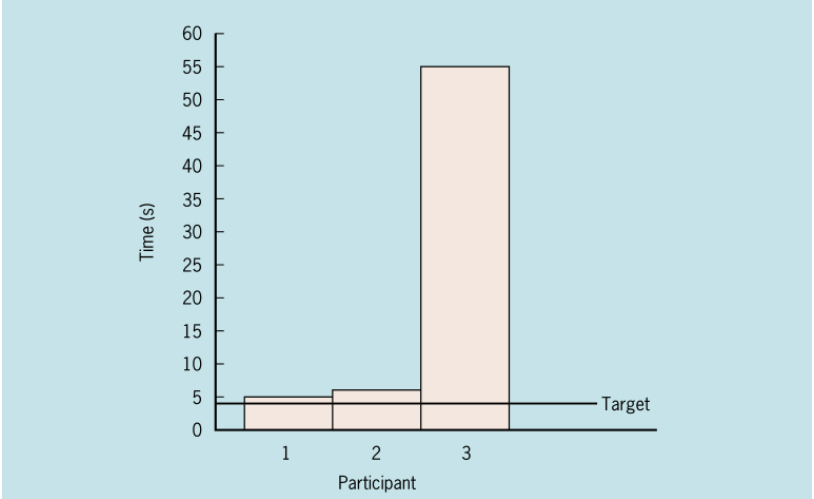
Summarizing Quantitative Data

- Tabular and visual representations

Example of Task Times

The task time for participant one (experienced) was five seconds. Participant two (experienced) took six seconds. The third participant was a novice and took 55 seconds. The target for task time is four seconds.

Participant	Experience level	Task time
Participant 1	Experienced	5 seconds
Participant 2	Experienced	6 seconds
Participant 3	Novice	55 seconds
Target		4 seconds



Recommendations

- After interpreting the results, you can produce recommendations, such as:
  - Successes to build on
  - Defects to fix
  - Possible defects or successes, without enough evidence (require further evaluation)
  - Areas of the UI not tested
  - Changes to usability and other requirements

Expert Reviews

- Often, designers make informal evaluations asking colleagues and clients for their opinion
- A more effective technique involves having an expert evaluator to test the design
  - staff or consultant
- The result of an expert evaluation can be:
  - a formal report with the list of problems identified and recommendations
  - a presentation or discussion with designers or managers

Expert Reviews

- Advantages:
  - The results are available quicker
  - Less expensive than user observation
  - Inspectors can suggest solutions to the defects
  - Help to find obvious errors easily, so they don't arrive to the evaluation with users
- Disadvantages:
  - Inspectors are not real users, and their prediction of what users will do with the UI or how important a defect is may fail
  - Inspectors have their own preferences about UI design, which may bias the evaluation data
  - Expert review is highly dependent on the inspector's experience

Expert Reviews

- The reviewer should
  - replicate the conditions of the end user (receive training on the system, read manuals, and use the system in a similar environment),
  - be sensitive in his/her recommendations,
  - realize that it is difficult for someone not directly involved to understand the design decisions and the development history,
  - have experience on the type of application being evaluated,
  - leave the development of the solutions to the developers,
  - be comprehensive in his/her report, and
  - review the consistency across all the windows of the application.

Expert Reviews

Evaluation methods

- Heuristic evaluation
  - The reviewer assess whether the interface follows a list of design heuristics (e.g., the Nielsen principles)
- Guidelines reviews
  - The expert checks that the interface follows the guidelines
- Standards inspection
  - Check that the design adheres to a standard (e.g. ISO 9241)
- Consistency inspection
  - Check terminology, fonts, color schemes, I/O formats, etc. within the interfaces as well as the documentation

Expert Reviews

Evaluation methods

BOX 4.1

Heuristics for the gaming environment.

1. Provide consistent responses to user's actions.
2. Allow users to customize video and audio setting, difficulty, and game speed.
3. Provide predictable and reasonable behavior for computer controlled units.
4. Provide unobstructed views that are appropriate for the user's current actions.
5. Allow users to skip non-playable and frequently repeated content.
6. Provide intuitive and customizable input mappings.
7. Provide controls that are easy to manage and that have an appropriate level of sensitivity and responsiveness.
8. Provide users with information on game status.
9. Provide instructions, training, and help.
10. Provide visual representations that are easy to interpret and that minimize the need for micromanagement.

From Pinelle et al., 2008.

Source: Designing the User Interface: Pearson New International Edition: Strategies for Effective Human-Computer Interaction. *Shneiderman, Ben* | *Cohen, Maxine; Plaisant, Catherine* ISBN 1-292-03701-6

Expert Reviews

Report

- The final report should
  - use guidelines to provide a good structure,
  - separate the problems depending on the users they affect,
  - rank the recommendations by their importance,
  - describe each recommendation at the conceptual level,
  - be aware of business and technical constraints,
  - solve the whole problem, not just a special case,
  - give specific and clear recommendations, with examples,
  - include small details like typos, poorly aligned data-entry fields, inconsistent layout of the controls, etc.

Expert Reviews

Report

- The final report should

BOX 4.2

Making usability recommendations useful and usable.

- Communicate each recommendation clearly at the conceptual level.
- Ensure that the recommendation improves the overall usability of the application.
- Be aware of the business or technical constraints.
- Show respect for the product team's constraints.
- Solve the whole problem, not just a special case.
- Make recommendations specific and clear.
- Avoid vagueness by including specific examples in your recommendations.

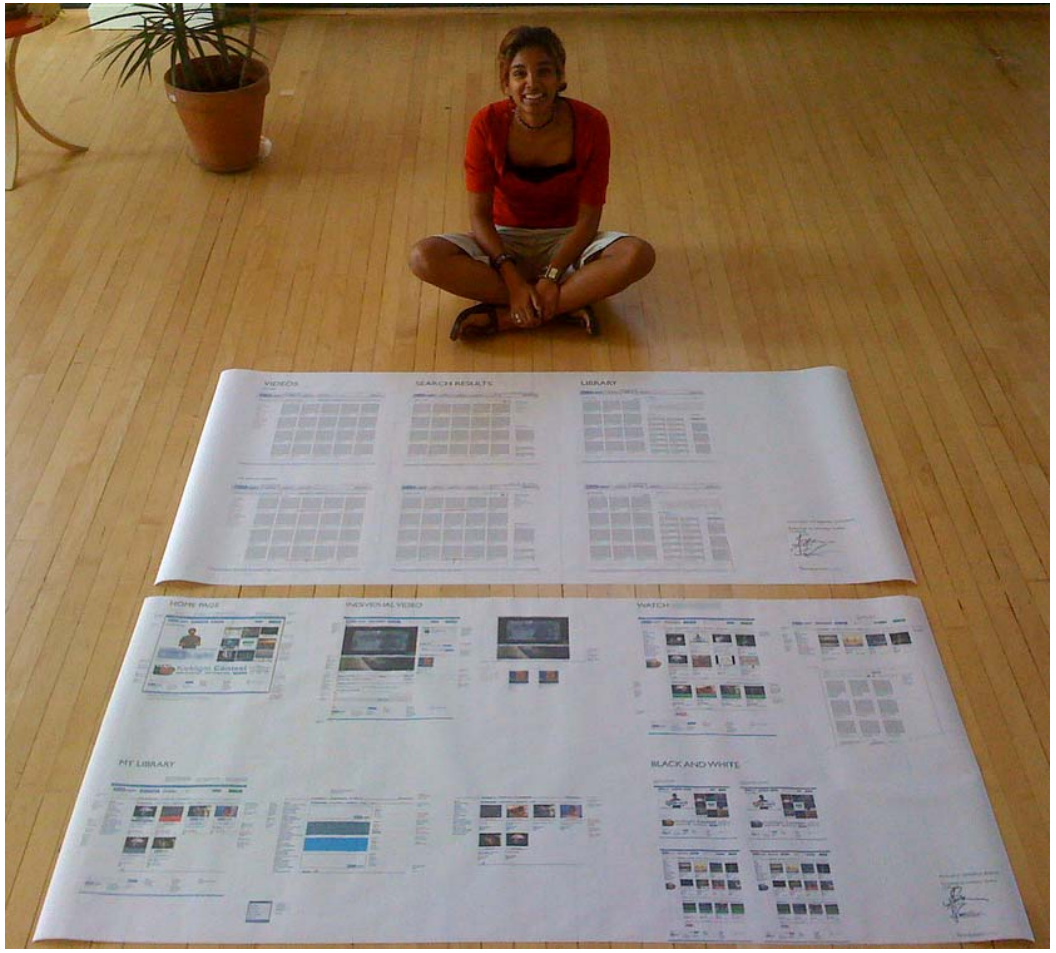
From Molich et al., 2007.

Source: Designing the User Interface: Pearson New International Edition: Strategies for Effective Human-Computer Interaction. *Shneiderman, Ben* | *Cohen, Maxine; Plaisant, Catherine* ISBN 1-292-03701-6

Expert Reviews

Other techniques

- Bird's-eye view
  - Study a full set of printed screens laid out on the floor or pinned to walls.
  - Detects inconsistencies (fonts, colors, terminology, similar actions)
  - Detects unusual patterns



flickr.com/photos/juhansonin/

Other Types of Evaluation

- Focus groups
  - About 8 people lead by a moderator discuss something
  - Widely used in marketing
  - Useful when the UI does not exists yet
- Card sorting
  - Ask users to group items
  - Useful to find underlying categories and structures
  - Print each item in a card and ask users to make groups, and probably give the group a name
  - Study the results, both:
    - where there are general agreement, and
    - where there is not agreement, and study how to solve it (maybe rename the item, or put the item in several categories)

Other Types of Evaluation

- Automatic validation tools
  - Accessibility checkers and HTML Validators
    - W3C Web Accessibility Initiative list: http://www.w3.org/WAI/ER/tools/
    - W3C CSS Validation Service: http://jigsaw.w3.org/css-validator/
    - SSB BART Accessibility Management Platform (AMP) http://www.ssbbartgroup.com/amp
    - …

Acceptance Tests

- Tests performed by the client to check that the delivered system meet the requirements
  - For example, establishing some test cases, with the response-time requirement of the hardware-software combination
- It is necessary to establish objective, measureable criteria:
  - Time for users to learn specific functions
  - Speed of task performance
  - Rate of errors by users
  - User retention of commands over time
  - Subjective user satisfaction

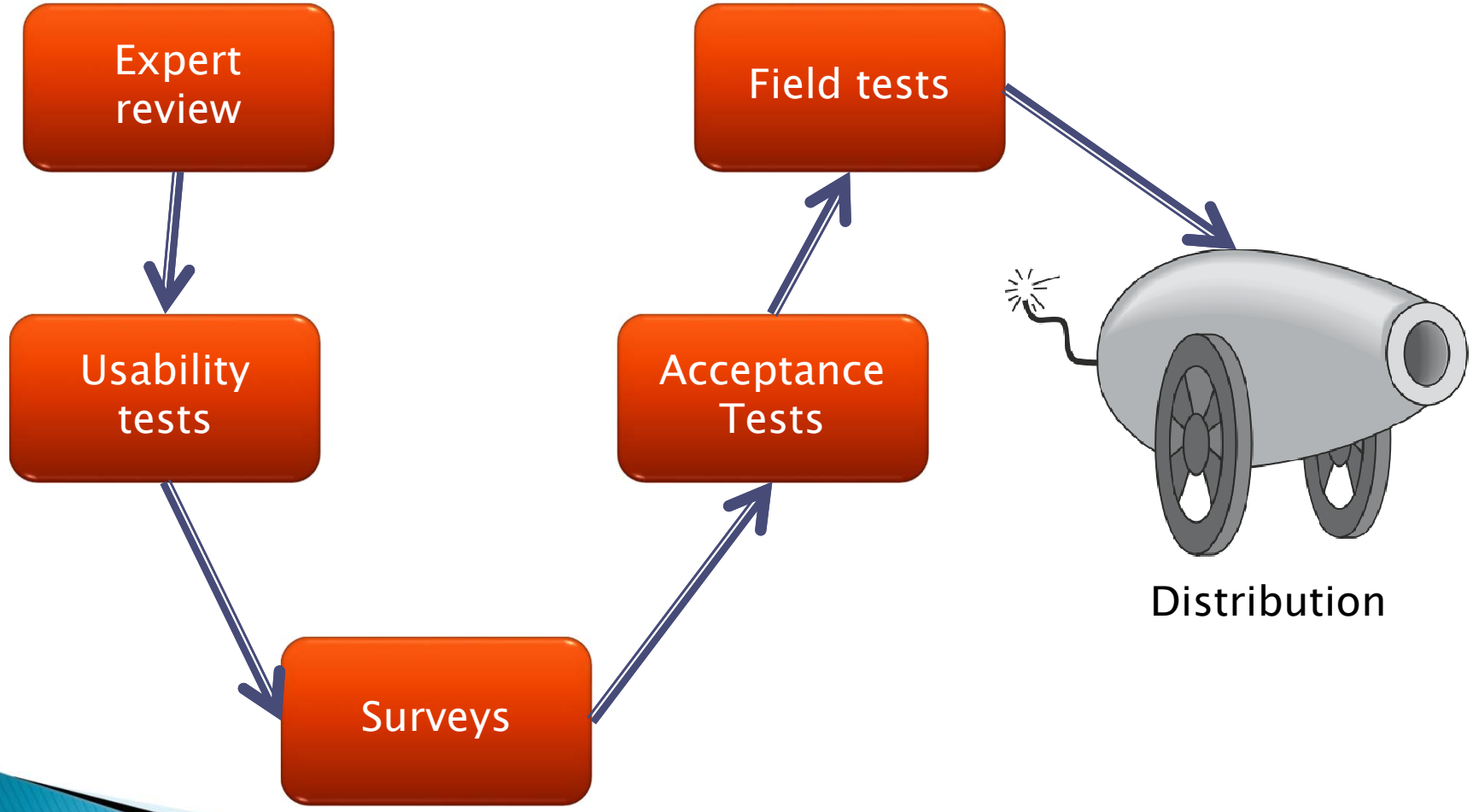
Acceptance Tests

- Example of an acceptance test for a food-shopping web site:
  - The participants will be 35 adults (25-45 years old), native speakers with no disabilities, hired from an employment agency. They will have moderate web-use experience: 1-5 hours/week for at least a year. They will be given a 5-minute demonstration on the basic features. At least 30 of the 35 adults should be able to complete the benchmark tasks within 30 minutes.

Acceptance Tests

- Other measurable parameters:
  - output comprehensibility, system response time, installation procedures, printed documentation, graphics appeal, etc.
- Precise acceptance criteria saves arguments and can demonstrate contractual fulfillment objectively
- Acceptance tests should be carried out by an outside organization
- After validation testing, there may be a period of field testing before distribution

Testing before distribution



Evaluation During Active Use

- After the system has been released, developers should study how it is being used to improve it
- As user numbers grow, major changes to the interface should be limited to an announced annual or semiannual revision
- Tools
  - Interviews and focus-group discussions
  - Continuous user-performance data logging
  - Online or telephone consultants, e-mail, and online suggestion boxes
  - Discussion groups, wikis, and newsgroups
  - Tools for automated evaluation

References

- D. Stone, C. Jarrett, M. Woodroffe. User Interface Design and Evaluation. Morgan Kaufmann, 2005.
  - Chapters 20-27
- Shneiderman, B. y Plaisant, C. Designing the User Interface. Pearson 5th ed., 2010
  - Chapter 4