

Pere-Pau Vázquez

IDI – Usability testing

Usability Testing. Outline

- Concepts
- Usability testing
- Usability laboratories
- Test roles
- Test planning



Pere-Pau Vázquez – pere.pau@cs.upc.edu



Usability. Concepts

- Usability:
 - Ease of use and acceptability of a system or product for a particular class of users carrying out specific tasks in a specific environment.
 - Where “ease of use” affects user performance, satisfaction.
 - And “Acceptability” affects whether or not the product is used.



Pere-Pau Vázquez – pere.pau@cs.upc.edu



Usability. Concepts

- Usability:
 - The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use.
 - To be useful, usability has to be specific. It must refer to particular tasks, particular environments and particular users.
 - So has to be its testing



Pere-Pau Vázquez – pere.pau@cs.upc.edu



Usability. Concepts

- How to test?
 - Ease of use is inversely proportional to the number and severity of difficulties people have in using software.
 - Let's examine the difficulties!!!



Pere-Pau Vázquez – pere.pau@cs.upc.edu



Usability. Usability testing

- Methods to evaluate usability:
 - Formal Usability tests (more on this later)
 - Interviews (typical from UCD)
 - Do not watch participants work
 - Usability inspection
 - Expert & heuristic evaluation
 - Do not observe participants performing work-like tasks
 - Field studies
 - Observe users under uncontrolled environments



Pere-Pau Vázquez – pere.pau@cs.upc.edu



Usability. Usability testing

- Methods to evaluate usability (ii)
 - Remote testing:
 - Lets participants with special needs, from other countries... to participate
 - May introduce familiar environments
 - May be difficult to have enough visual feedback from the participant
 - May lead to compromised security of unpublished products



Pere-Pau Vázquez – pere.pau@cs.upc.edu



Usability. Usability testing

- Think Aloud:
 - Can be part of almost any testing technique
 - Participants must talk about what they are doing as they do it
 - Prompt participants to resume if they stop talking
 - What users say during tasks is more reliable than posterior interviews
 - Seem to work better with pairs of participants
 - Better for problem discovery than measurement



Pere-Pau Vázquez – pere.pau@cs.upc.edu



Usability. Usability testing

- Formal usability tests:
 - Can be very informal or very formal
 - Observer might sit next to the participant, watch through a one-way glass, or watch the on-screen behaviour of a participant who is performing specified tasks
 - Often use think-aloud (TA)
 - Observers might watch one or two participants at a time
 - Evaluated software can be varied:
 - Prototypes, under development, competitive products...



Pere-Pau Vázquez – pere.pau@cs.upc.edu



Usability. Usability testing

- Two major families by goals:
 - Problem discovery: Discovery, prioritization, and resolution of usability problems
 - May be informal
 - Measurement: Include two fundamental tasks:
 - The development of the usability objectives.
 - Iterative testing to determine if the product under test has met the objectives



Pere-Pau Vázquez – pere.pau@cs.upc.edu



Usability. Usability laboratories

- Set of soundproofed rooms
 - Participant area
 - Observer area with one-way glass
 - Video cameras and microphones
 - Executive viewing area behind the primary observer area



Pere-Pau Vázquez – pere.pau@cs.upc.edu



Usability. Test roles

- Test administrator: Designs the usability study
 - Specifies the initial conditions for the test session and the codes to use for data logging.
 - Conducts reviews with the rest of the test team
 - Leads the data analysis
 - Puts together the final presentation or report



Pere-Pau Vázquez – pere.pau@cs.upc.edu



Usability. Test roles

- Briefer: Interacts with participants
 - Briefs participants at the start of the test
 - Communicates with them as required during the test
 - Debriefs participants at the end of the test sessions
 - In a think-aloud study, the briefer has the responsibility to keep the participant talking
 - Needs to be familiar enough with the product in order to decide what to answer to the participants



Pere-Pau Vázquez – pere.pau@cs.upc.edu



Usability. Test roles

- Camera Operator: Responsible for running the audio-visual equipment during the test
- Data Recorder: Writes is the notes that the data recorder takes during a test session.
 - Is the primary data used for the usability study
 - The camera may not catch the important action at every moment of a usability study
 - Usually uses data-logging software
 - It is a very demanding skill



Pere-Pau Vázquez – pere.pau@cs.upc.edu



Usability. Test roles

- Help Desk Operator: Replaces a real help desk operator
 - Required when the participant experiences enough difficulty to place a call
 - Must behave as a call-center person
- Product Expert: Maintains the product and offers technical guidance during the test
 - Must recover if there is a product failures
 - Helps the other team members understand the system's actions during the test
- Statistician: Extracts the maximum amount of information from the data gathered during a formal test
 - Rarely required for informal tests



Pere-Pau Vázquez – pere.pau@cs.upc.edu



Usability. Test planning

- Before Starting, the administrator must:
 1. Understand the purpose of the product
 2. Decide which parts of the product are ready for testing
 3. Determine the types of people who will use the product
 4. Determine the use given to the product
 5. Determine the conditions of usage of the product



Pere-Pau Vázquez – pere.pau@cs.upc.edu



Usability. Test planning

- Determining the purpose of the test
 - Measurement vs Usability problems identification
 - Product comparison
 - Within-subjects or between-subjects
 - Statistical analyses may vary
- Participants
 - Develop user profile
 - Sometimes available from the product's documentation
 - Participants must be representative of end users



Pere-Pau Vázquez – pere.pau@cs.upc.edu



Usability. Test Planning: Measures

- For problem discovery:
 - Prioritize problems
 - Include frequency of occurrence
 - Likelihood of occurrence in normal use
 - Magnitude of impact
 - Pre-planned number of iterations



Pere-Pau Vázquez – pere.pau@cs.upc.edu



Usability. Test Planning: Measures

- For measurement tests:
 - Categories
 - Goal achievement indicators (success rate and accuracy)
 - Work rate indicators (speed and efficiency)
 - Operability indicators (error rate and function usage)
 - Knowledge acquisition indicators (learnability and learning rate)



Pere-Pau Vázquez – pere.pau@cs.upc.edu



Usability. Test Planning: Measures

- For measurement tests:
 - Measures
 - Successful task completion rates
 - Mean task completion times
 - Mean participant satisfaction ratings (on a task-by-task basis)
 - There are standardized questionnaires for this
 - Other measurements could be:
 - Number of tasks completed within a specified time limit, number of wrong menu choices, number of user errors, number of repeated errors (same user)



Pere-Pau Vázquez – pere.pau@cs.upc.edu



Usability. Test Planning

- After measurements choice, goals can be determined
 - It's usually better to set goals that make reference to an average (mean) than to a percentile
 - Sample means drawn from a continuous distribution are less variable than sample medians
 - Unless there is missing data due to participants failing to complete tasks
 - Percentile goals require large sample sizes
 - You can't measure accurately at the 95 percentile unless there are at least twenty measurements



Pere-Pau Vázquez – pere.pau@cs.upc.edu



Usability. Test Planning

- Goal determination
 - For counting events (i. e. successful task completion) use percentiles
 - Unless lower than roughly 100% rates are acceptable



Pere-Pau Vázquez – pere.pau@cs.upc.edu



Usability. Testing

- Participants
 - User profile must be developed
 - Sometimes available from the marketing group
 - Can be obtained from employment agencies, internal sources, market research firms, existing customers...
 - Must define the characteristics of the target population
 - They are difficult to define:
 - May involve previous experience, education level, age, sex...



Pere-Pau Vázquez – pere.pau@cs.upc.edu



Usability. Testing

- Participants. Factors to consider:
 - Demographic locations
 - Age ranges
 - Levels of experience
 - Levels of gender
- Number of users:
 - Will depend on many factors
 - Money and time
 - Type of study: Single-shot (larger) vs iterative (smaller)



Pere-Pau Vázquez – pere.pau@cs.upc.edu



Usability. Testing

- Test task scenarios:
 - Must be representative
 - Core tasks: Features that everybody uses
 - Peripheral tasks: Features used less often
 - Once the tasks are defined, scenarios of use must be created
 - Define initial conditions
 - Description of the scenario: what to do and why
 - Some action must be taken on finish
 - Should not provide step-by-step instructions but should include details
 - Not all users must be provided with the same scenarios (may depend on the user profile)



Pere-Pau Vázquez – pere.pau@cs.upc.edu



Usability. Testing

- Procedure:
 - Introduction: Purpose of the test, confidentiality...
 - Task performance:
 - Complete preliminary questions and forms (background questionnaire, informed consent form, confidential disclosure form...)
 - Complete training (if required)
 - Perform the tasks



Pere-Pau Vázquez – pere.pau@cs.upc.edu



Usability. Testing

- Procedure:
 - Task performance:
 - Usually, no help is provided:
 - Refer the users to the documentation
 - If required, provide help, but score the task as failed
 - Try to avoid direct answers to questions
 - If asking questions, try to avoid biasing the participant's response
 - Give a satisfaction questionnaire at the end of each scenario.
 - After the scenarios, final questionnaire
 - There are standardized versions



Pere-Pau Vázquez – pere.pau@cs.upc.edu



Usability. Testing

- Pilot testing:
 - Usability test must be tested
 - Commonly, a member of the usability team can do the testing.



Pere-Pau Vázquez – pere.pau@cs.upc.edu



Usability. Testing

- Reporting:
 - Describe usability problems
 - Present quantitative measurements



Pere-Pau Vázquez – pere.pau@cs.upc.edu



Usability. Testing

- Reporting. Usability problems:
 - Should lead to a recommendation
 - Should indicate the severity
 - Can be classified:
 - Mistakes: Errors due to incorrect intention
 - Slips: Errors due to appropriate intention but incorrect action
 - Expertise does not affect on the number of errors
 - But affects how fast they are handled



Pere-Pau Vázquez – pere.pau@cs.upc.edu



Usability. Testing

- Reporting. Recommendations:
 - Global changes first
 - Must be checked:
 - A missing help may be a global problem or something related with a concrete UI
 - Try to give at least one recommendation for each problem
 - Present the different trade-offs clearly



Pere-Pau Vázquez – pere.pau@cs.upc.edu



Usability. Testing

- Reporting. Prioritizing problems:
 - Judgement driven
 - Data driven: frequency, impact, ease of correction...



Pere-Pau Vázquez – pere.pau@cs.upc.edu



Usability. Testing

- Reporting. Quantitative measurements:
 - Provide means, standard deviations, and confidence intervals
 - Common problems: Failure to meet targets, large standard deviation



Pere-Pau Vázquez – pere.pau@cs.upc.edu



Usability. Testing

- Sample size estimation:
 - [Virzi, 1992] found that 80% of known usability problems could be surfaced with 5 testers, and 3 that testers would reveal the most severe problems
 - [Nielsen & Landauer, 1993] say that the best benefits are usually obtained testing no more than 5 users and running as many small tests as you can afford



Pere-Pau Vázquez – pere.pau@cs.upc.edu



Usability. Testing

- Sample size estimation:
 - Quality of tests may seriously affect the number of detected problems [Faulkner, 2003]
 - Quality of testers also has an impact on the number of usability problems revealed [Faulkner, 2003]



Pere-Pau Vázquez – pere.pau@cs.upc.edu



Usability. Testing

- Sample size estimation:
 - There is a law of diminishing returns [Nielsen, 2000]
 - The third tester will do many things that you have already observed with the first or second user
 - Will generate a small amount of new data
 - After the fifth user you are wasting your time by observing the same findings repeatedly but not learning much new



Pere-Pau Vázquez – pere.pau@cs.upc.edu



Usability. Testing

- Safe values:
 - 3-4 users to find main problems (≈70-80%)
 - 5-6 users to find most problems



Pere-Pau Vázquez – pere.pau@ics.upc.edu



Pere-Pau Vázquez

IDI – Usability