Evaluating Interface Designs

- Extensive testing and evaluation are necessities.

- Expert reviews and Heuristics

- A natural starting point for evaluating new or revised interfaces is to present them to colleagues or customers and ask for their opinions.
- Such informal demos with test subjects can provide some useful feedback,
- More formal *expert reviews* have prove to be far more effective.
- These methods depend on having experts (whose expertise may be in the application or user-interface domain) available on staff or as consultants.
- The reviews can then be conducted rapidly and on short notice by having the expert walk through the key functionality of the interface using a disciplined approach.

- Expert reviews can occur early or late in the design phase.
- The outcome may be a formal report with problems identified or recommendations for changes.
- Alternatively, the expert review may culminate in a discussion with or presentation to designers or managers. .
- Expert reviews usually take from half a day to one week..
- **Heuristic evaluation**. The expert reviewers critique an

- The Expert reviews can be conducted rapidly and on short notice by having the expert walk through the key functionality of the interface using a disciplined approach.
- Expert reviews can occur early or late in the design phase.
- The outcome may be a formal report with problems identified or recommendations for changes.
- Alternatively, the expert review may culminate in a discussion with or presentation to designers or
- managers.

Heuristic Evaluation

- . The expert reviewers critique an interface to determine conformance with a short list of design heuristics, such as the Eight Golden Rules ..
- Although interfaces have changed vastly over the years, the creation of most sets of heuristics is based on these eight golden rules.

Heuristics for the gaming environment

- Provide consistent responses to user's actions.
- Allow users to customize video and audio setting, difficulty, and game speed.
- Provide predictable and reasonable behavior for computer controlled units.
- Provide unobstructed views that are appropriate for the user's current actions.
- Allow users to skip non-playable and frequently repeated content.
- Provide intuitive and customizable input mappings.
- Provide controls that are easy to manage and that have an appropriate

- appropriate level of sensitivity and responsiveness.
- Provide users with information on game status.
- Provide instructions, training, and help.
- Provide visual representations that are easy to interpret and that minimize the need for micromanagement

- **Guidelines review**. The interface is checked for conformance with the organizational or other guidelines document .
- Because guidelines documents may contain a thousand items or more, it may take the expert reviewers some time to absorb them and days or weeks to review a large interface.

- **Consistency inspection**. The experts verify consistency across a family of interfaces, checking the terminology, fonts, color schemes, layout, input and output formats, and so on, within the interfaces as well as any supporting materials.
- Software tools can help automate the process as well as produce concordances of words and abbreviations.
- Often large-scale interfaces may be developed by several groups of designers; this can help smooth over the interface and provide a common and consistent look and feel.
- **Cognitive walkthrough**. The experts simulate users walking through the interface to carry out typical tasks.
- **Formal usability inspection**. The experts hold a courtroom-style meeting, with a moderator or judge, to present the interface and to discuss its merits and weaknesses.
- Design-team members may rebut the evidence about problems in an adversarial format.
- Formal usability inspections can be educational experiences for novice designers and managers, but they may take longer to prepare and need more personnel to carry out than do other types of review

Usability Testing

- usability tests are designed to find flaws in user interfaces.

Types of usability Testing

- **Paper mockups and prototyping**. Early usability studies can be conducted using paper mockups of pages to assess user reactions to wording, layout, and sequencing.
- A test administrator plays the role of the computer by flipping the pages while asking a participant user to carry out typical tasks.
- This informal testing is inexpensive, rapid, and usually productive.
- Typically designers create *low-fidelity* paper prototypes of the design, but today there are computer programs (e.g., Microsoft Visio, SmartDraw, Gliffy, Balsamiq, MockingBird) that
- can allow designers to create more detailed *high-fidelity* prototypes
- with minimal effort. Interestingly enough, users have been shown to
- respond more openly to the lower-fidelity designs, potentially

- **Discount usability testing**. Discount usability testing is used as a *formative evaluation* (while designs are changing substantially) and more extensive usability testing as a *summative evaluation* (near the end of the design process).
- The formative evaluation identifies problems that guide redesign
- summative evaluation provides evidence for product announcements ("94% of our 120 testers completed their shopping tasks without assistance") and clarifies training needs ("with four minutes of instruction, every participant successfully

3

- **Competitive usability testing**. Competitive testing compares a new interface to previous versions or to similar products from competitors.
- This approach is close to a controlled experimental study and staff must be careful to construct parallel sets of tasks and to counterbalance the order of presentation of the interfaces.
- .

- **A/B testing.** This method tests different designs of an interface.
- Typically, it is done with just two groups of users to observe and record differences between the designs.
- It is also referred to as *bucket testing*, it is similar to a *between-subjects design* . This method of testing is often used with large-scale online controlled experiments
- A/B testing involves randomly assigning two groups of users to either the control group (no change) or the treatment group (with the change) and then having some dependent measure that can be tested to see if there is a difference between the groups .
- Before running A/B testing, it is often suggested to run an *A/A test* or a null test.
- In A/A testing, there are still two groups, but they both receive the same treatment (the control); then the variability for power calculations and the experimentation system can be tested.
- In a true test, with a 95% confidence level, the null hypotheses should be rejected.
- This testing method has been used at Bing, where more than 200 experiments are run concurrently with 100 million customers spanning billions of changes.

5

- **Universal usability testing**. This approach tests interfaces with highly diverse users, hardware, software platforms, and networks.
- When a wide range of international users is anticipated, such as for consumer electronics products, web-based information services, or e-government services, ambitious testing is necessary to clean up problems and thereby help ensure success.
- Trials with small and large displays, slow and fast networks, and a range of operating systems or Internet browsers will do much to improve the user experience.

6

- **Field tests and portable labs**. This testing method puts new interfaces to work in realistic environments or in a more naturalistic environment in the field for a fixed trial period.
- These same tests can be repeated over longer time periods to do longitudinal testing.
- Field tests can be made more fruitful if logging software is used to capture error, command, and help frequencies as well as productivity measures.
- Portable usability laboratories with recording and logging facilities have been developed to support more thorough field testing.

7

- **Remote usability testing.** Since web-based applications are available across the world, it is tempting to conduct usability tests online, avoiding the complexity and cost of bringing participants to a lab.
- This makes it possible to have larger numbers of participants with more diverse backgrounds, and it may add to the realism, since participants do their tests in their own environments and use their own equipment.
- Participants can be recruited by e-mail from customer lists or through online communities including Amazon Mechanical Turk.
- This opens the pool of participants to sophisticated users who, - perhaps because of their remote locations or other physical challenges, could not otherwise get to a lab location.
- The downside is that there is less control over user behaviors and diminished ability to observe users' reactions, although usage logs and phone interviews are useful supplements.
- These tests can be performed both synchronously (users do tasks at the same time while the evaluator observes) and asynchronously (users perform tasks independently and the evaluator looks at the results later).
- Some studies have shown remote usability testing to find more problems than traditional usability testing.

- **Can-you-break-this tests**. Game designers pioneered the can-youbreak- this approach to usability testing by providing energetic teenagers with the challenge of trying to beat new games.
- This destructive testing approach, in which the users try to find fatal flaws in the system or otherwise destroy it, has been used in other types of projects as well and should be considered seriously.
- Users today have little patience with flawed and poorly designed products and are often fickle with company loyalty if reasonable competitors exist

Acceptance Tests

- For large implementation projects, the customer or manager usually sets objective and measurable goals for hardware and software performance.
- In the requirements documents the mean time between failures as well as the mean time to repair for hardware and, in some cases, software Failures may be specified.
- Explicit acceptance criteria should be established when the requirements document is written or when a contract is offered.
- Measurable criteria for the user interface can be established for the following:
- **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**
- After successful acceptance testing, there may be a period of field testing or an extensive beta test with real users before national orinternational distribution.
- In addition to further refining the user interface, field tests can improve training methods, tutorial materials, telephone-help procedures, marketing methods, and publicity strategies

- The goal of early expert reviews, usability testing, surveys, acceptance testing, and field testing is to force as much as possible of the evolutionary development into the pre-release phase, when change is relatively easy and less expensive to accomplish.