Interface Evaluation

- **Evaluation Techniques**
  - Test usability and functionality, occurs in lab, field or in collaboration with users
  - Can evaluate both design and implementation, should be considered at all stages
- **Determine the goals**
  - Goals lead the approach taken for the study
  - Examples: identify best metaphor to base design on, check that final interface is consistent, investigate how technology affects working practices, improve usability of existing product
- **Determine the questions**
  - Need goals and questions to guide evaluations
  - Examples: finding out why customers prefer to purchase paper tickets instead of e-tickets can become ‘sub-questions’:
    - What are the attitudes to the new tickets, are they concerned about them being accepted, is the interface for obtaining them poor
- **Choose the approach and methods**
  - Approach influences the methods used, and in turn, how data is collected, analysed and presented
  - E.g. field study typically:
    - Involves observation and interviews, does not involve lab tests, produces qualitative data
  - **Practical issues**
    - Selecting users, staying on budget, on schedule, finding evaluators, selecting equipment
  - **Ethical issues**
    - Develop a consent form
    - People have a right to: know the goals of the study, what will happen to the findings, privacy, leave at will, be treated politely
- **Evaluate interpret and present data**
  - Study should be: reliable (reproducible), valid, unbiased, well scoped, environment agnostic
- **Lab studies**
  - Advantages: specialist equipment available, uninterrupted environment
  - Disadvantages: lack of context, difficult to observe users cooperating
  - Appropriate: if field location would be dangerous, or impractical
- **Field Studies**
  - Advantages: natural environment, context retained, long period studies
  - Disadvantages: distractions, noise
  - Appropriate: if context is vital and requires long term study
- **Heuristic Evaluation**
  - Usability criteria (heuristics) are identified, design examined by experts to see if they are violated
  - Example heuristics: system behaviour is predictable, consistent, and provides feedback

    - **Nielsen's heuristics**
        - **Visibility of system status**
        - **Match between system and real world** - follows real world conventions
        - **User control and freedom** - support undo and redo, allow exploring
        - **Consistency and standards**
        - **Error Prevention**
        - **Recognition rather than recall** - make options visible
        - **Flexibility and efficiency of use** - cater to novice and experts
        - **Aesthetic and minimalist design** - important info only
        - **Help users recognize, diagnose, and recover from errors**
        - **Help and documentation**
- **Guidelines for tabletop displays**
    - Support interpersonal interaction, fluid transition between activities, use of physical objects, simultaneous actions
- **Evaluating implementations**
    - Requires an artefact - e.g. simulation, prototype, full implementation
    - **Experimental evaluation** - controlled evaluation of specific behaviour, choose hypothesis to be tested, conditions only differ based on a controlled variable
    - **Experimental factors**
        - Subjects - representative, sufficient sample
        - Variables
            - Independent variable (e.g. interface style, number of menu items) - what we change
            - Dependent variable (e.g. time taken, number of errors) - the result
        - Hypothesis: prediction of outcome based on IV and DV
            - Null hypothesis states no difference will be seen
        - Within subjects (repeated measures) design
            - Each subject evaluated in each condition, less costly, users may learn as they go
        - Between subjects (randomised) design
            - Each subject performs under a single condition, no learning, more users needed, variation can cause bias
- **Data Analysis**
    - Before statistics - check data purity
    - Choose technique based on type of data and information required
        - Discrete/Continuous
    - Contingency table - classify data by several discrete attributes, count items per group
- **Observation methods - think aloud**
    - User observed performing task, asked to describe their actions, motivations, what they think is happening
    - Advantages: simplicity, provides useful insight, shows how system is actually used

    - Disadvantages: subjective, selective, act of describing may alter task performance
    - Can get very valuable info: what is the user trying to do vs what they are doing
    - **Procedure**
        - Ensure user knows the system is being judged, not them
        - Get them to comment liberally, give minimal help, help given should be recorded
        - Observer takes notes about what is being done/said
        - **Two approaches:** observer specifies specific task, or user is free to interact and choose their own task
    - **Prompts:** what are you thinking now; why did you do that
- **Observation methods - cooperative evaluation**
    - Variation on think aloud
    - User and evaluator ask each other questions throughout
    - Advantages: less constrained, encourages user to criticise the system, can give clarification
- **Protocol Analysis**
    - Record of evaluation session is known as protocol
    - Paper and pencil - cheap, limited to writing speed
    - Audio recording - good for thinking aloud, hard to sync with other protocols
    - Video recording - accurate and realistic, needs special equipment, obtrusive
    - Computer logging - automatic and unobtrusive, large data volume, difficult to analyse
    - Post task walkthrough: user explains decisions after the event, provides intention
- **Query Techniques**
    - **Interviews**
        - Analyst questions user on one to one basis - informal, subjective and cheap
        - Advantages: varied to suit context, explore issues, find unanticipated issues
        - Disadvantages: subjective and time consuming
    - **Questionnaires**
        - Fixed questions given to users
        - Advantages: quick, large spread, analysed more rigorously
        - Disadvantages: less flexible, less probing
        - Requires careful design, and cannot have leading questions
        - Question types: general, open-ended, scalar, multi-choice, ranked
- **Choosing an Evaluation method**
    - When is the evaluation being carried out - design vs implementation
    - What style of evaluation is required - lab vs field
    - How objective should it be - objective vs subjective
    - Type of measures required - qualitative vs quantitative
    - Level of information required - high level vs low level
    - Level of interference required - obtrusive vs unobtrusive

    - Resources available - time, subjects, equipment, expertise
- **Evaluating User Experience**
    - Qualitative measures: how easy was it, effectiveness, enjoyable, types of errors
    - Quantitative measures: time taken, number of errors
- **Trust**
    - Extent of a person's confidence in the recommendations and decisions of an artificially intelligent aid
    - **Scale**
        - **Perceived Reliability** - reliable performance, useful advice
        - **Perceived Technical Competence** - appropriate decision making approach, sound knowledge base
        - **Perceived Understandability** - understand how system behaves
        - **Faith** - belief in system advice
        - **Personal Attachment** - sense of loss in system absence