Avaliação de Interfaces

Interface Pessoa-Máquina - 25/26 - LEI / UM

Hugo Pacheco
hpacheco@di.uminho.pt

Software Quality

- What defines the quality of an (interactive) software system?
  - ISO/IEC 25000 — System and Software Quality Requirements and Evaluation (2014)
  - ISO/IEC 25010 — System and Software Quality Models (2011)

Usability Evaluation

- “**Usability** is 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” (ISO 9241-11)

- “**Evaluation** is a process that critically examines a program. It involves collecting and analyzing information about a program’s activities, characteristics, and outcomes. Its purpose is to make judgments about a program, to improve its effectiveness, and/or to inform programming decisions” [1]

- “Put simply, **usability evaluation** assesses the extent to which an interactive system is easy and pleasant to use” [2]

- Mas como medir usability?

[1] Patton, M.Q. (1987). Qualitative Research Evaluation Methods. Thousand Oaks, CA: Sage Publishers.
[2] http://www.interaction-design.org/encyclopedia/usability_evaluation.html

Evaluation Methods

- **Analytical methods**
  - Identificar potenciais problemas de usabilidade
  - Com peritos
  - Sem utilizadores
  - Baixo custo: tipicamente baseados em modelos ou protótipos do sistema
  - Também denominados métodos **preditivos**
- **Empirical methods**
  - Avaliar usabilidade e identificar problemas efectivos
  - Medir a realidade
  - Com utilizadores
  - Custo elevado: recrutar utilizadores, obter condições realistas, recolher e avaliar dados
  - Também denominados métodos **interpretativos**

Evaluation Methods

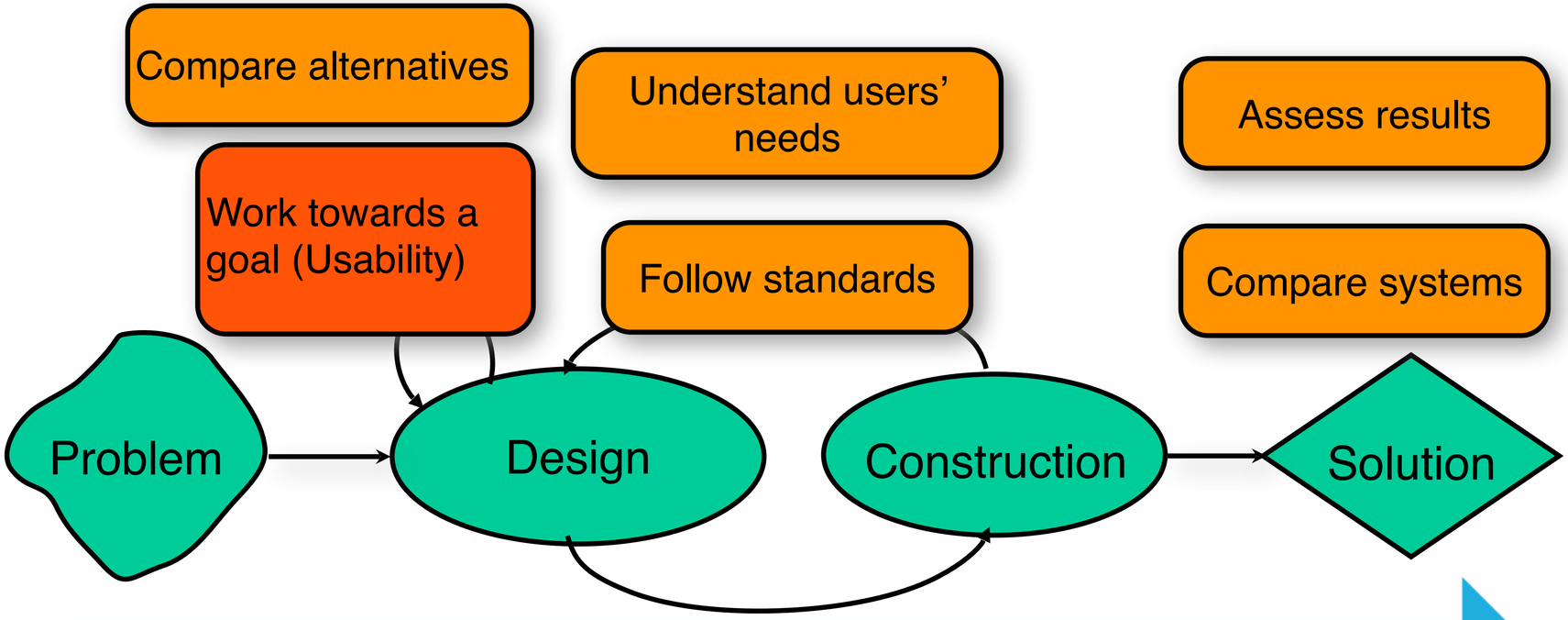
- “If you want to evaluate a tool, say an axe, you might study the design of the bit, the weight distribution, the steel alloy used, the grade of hickory in the handle, etc., or you may just study the kind and speed of the cuts it makes in the hands of a good axeman.”
- **Analytical evaluation** identifies the crucial characteristics
  - “If the axe does not cut well, what do we have to change?”
- **Empirical evaluation** helps to understand the context for object properties
  - “Why does the axe have a special-shaped handle?”

[1] Scriven, M. (2013). Conceptual revolutions in evaluation. Evaluation roots, SAGE Publications, 167-179

Assessment Methods

- **Formative evaluation**
  - How do you design it?
  - Improvement-oriented
  - Prototyping, sketching
  - Interviews, surveys, diaries, case studies
  - Focus groups, user role playing, personas
- **Summative evaluation**
  - How good is it?
  - Based on criteria
  - Depends on the development
  - Ratings, questionnaires, scales
  - Efficiency measures
  - Physiological measures

Assessment Methods



"What are the requirements of the system?"	*"Does the system operating as planned?"*	*"Do the users achieve (all of) their objectives?"*	*"What is the predicted and actual impact?"*
requirement analysis	prototype evaluation	program evaluation	impact evaluation

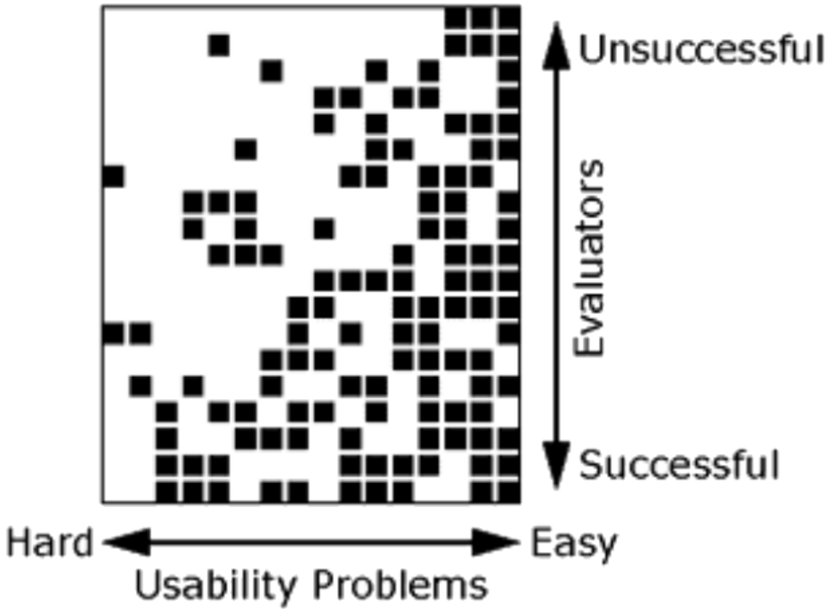
Usability Evaluation

- Várias abordagens, apenas algumas exemplificadas
- Embora
  - Analytical ⇝ Formative
  - Empirical ⇝ Summative
- Eixos de classificação são ortogonais

Methods	Formative	Summative
Analytic	▪ Cognitive walkthrough ▪ Usability Action Notation (UAN)	▪ Heuristic evaluation ▪ Standards compliance
Empirical	▪ User group identification ▪ Prototype user study	▪ Usability/UX testing ▪ Field studies

Heuristic Evaluation

- **Analytical & Summative**
- A “discount usability engineering method” for quick, cheap, and easy evaluation of a UI design
- The most popular usability evaluation method
- A small set of evaluators examines the interface and judge its compliance with recognized usability guidelines (the “heuristics")
- Goal: identifying any problems associated with the design of user interface
- Main question: is the interface compatible with the intended users' needs and preferences?



Nielsen, Jakob, and Rolf Molich. “Heuristic evaluation of user interfaces.” Proceedings of the SIGCHI conference on Human factors in computing systems. 1990.

Heuristic Evaluation

- Main advantages
  - ✚ May produce useful results with modest investment
  - ✚ Simple to apply even by not very experienced evaluators
  - ✚ May be used along the development process from early phases
- Main limitations
  - ➖ Subjective (partially overcome with more experienced evaluators)
  - ➖ Tends to find many small problems which may not be very important
  - ➖ Can’t find all usability problems

Cognitive Walkthrough

- **Analytical & Formative**
- One or more experts (e.g. design team) specify and perform a series of tasks, evaluating the design for each one. Requires:
    - Uma descrição de quem são os utilizadores
    - Descrições detalhadas das principais tarefas a serem realizadas no sistema
    - Um protótipo do sistema que permita simular a execução das tarefas
- Goal: assessing learnability, based on the fact that users usually prefer to learn a system by using it
- Main question: how easy it is for new users to accomplish tasks with the system?

Lewis, Clayton, et al. "Testing a walkthrough methodology for theory-based design of walk-up-and-use interfaces." *Proceedings of the SIGCHI conference on Human factors in computing systems*. 1990.

Cognitive Walkthrough: procedimento

- Cada perito coloca-se no lugar do utilizador e tenta **executar uma tarefa**

    1. Formula um **objectivo** a ser atingido no sistema
    2. Procura na interface as **acções** disponíveis
    3. Selecciona e realiza a acção que lhe parece mais **adequada** face ao seu objectivo
    4. **Avalia** a resposta do sistema para avaliar se está a fazer progresso

- Cada perito vai respondendo a um **conjunto pré-definido de questões**

    1. A acção correcta é suficientemente evidente para o utilizador?
    2. O controlo para executar a acção está visível?
    3. Irá o utilizador associar a acção correcta ao controlo?
    4. Irá o utilizador interpretar de forma correcta a resposta do sistema?

Cognitive Walkthrough: procedimento

- Respostas **positivas** ou…
- Respostas negativas devem ser classificadas por severidade:
  - **Tipo 1:** o problema pode causar confusão ou demora na execução da tarefa
  - **Tipo 2:** o problema pode impedir que o utilizador consiga executar a tarefa sem ajuda
  - **Tipo 3:** o problema impede a execução da tarefa
- Análise deve ser efectuada por 5-7 avaliadores e os resultados são agregados

Cognitive Walkthrough: exemplo

- Utilizador = visitante de hospital, acostumado a máquinas de vending “tradicionais”
- Tarefa a analisar
  1. Inserir moedas no moedeiro
  2. Seleccionar pod, premindo o botão correspondente
  3. Recolher pod e inseri-lo na ranhura
  4. Recolher troco, se necessário
  5. Pegar em copo da lateral e colocá-lo por baixo do bico
  6. Seleccionar tamanho de café premindo o botão correspondente
  7. Recolher copo com café

1. A acção correcta é suficientemente evidente para o utilizador?
2. O controlo para executar a acção está visível?
3. Irá o utilizador associar a acção correcta ao controlo?
4. Irá o utilizador interpretar de forma correcta a resposta do sistema à acção escolhida?

Cognitive Walkthrough: formulário

Sistema: _____
Tarefa: _____
Utiliza.: _____

Passos	Questões	OK? (S/ N)	Risco(1-3)	Problema /Sugestão
1	1			
	2			
	3			
	4			
2	1			
	2			
	3			
	4			
3	1			
	2			
	3			
	4			
4	1			
	2			
	3			
	4			
5	1			
	2			
	3			
	4			
6	1			
	2			
	3			
	4			
7	1			
	2			
	3			
	4			

1. A acção correcta é suficientemente evidente para o utilizador?
2. O controlo para executar a acção está visível?
3. Irá o utilizador associar a acção correcta ao controlo?
4. Irá o utilizador interpretar de forma correcta a resposta do sistema à acção escolhida?

Cognitive Walkthrough

- Advantages
  - ✚ Flexible, quick, and easy to do
  - ✚ Can be used in early development stages
  - ✚ Errors recognizable in the approach
  - ✚ Makes explicit the decisions that have been made in the process of designing an interface
- Disadvantages
  - ➖ Artificial setting and tasks
  - ➖ Not suitable for complex systems. Extensive and often very time consuming
  - ➖ Evaluator cannot objectively assess the user
  - ➖ Can’t find all usability problems

Analytical Methods

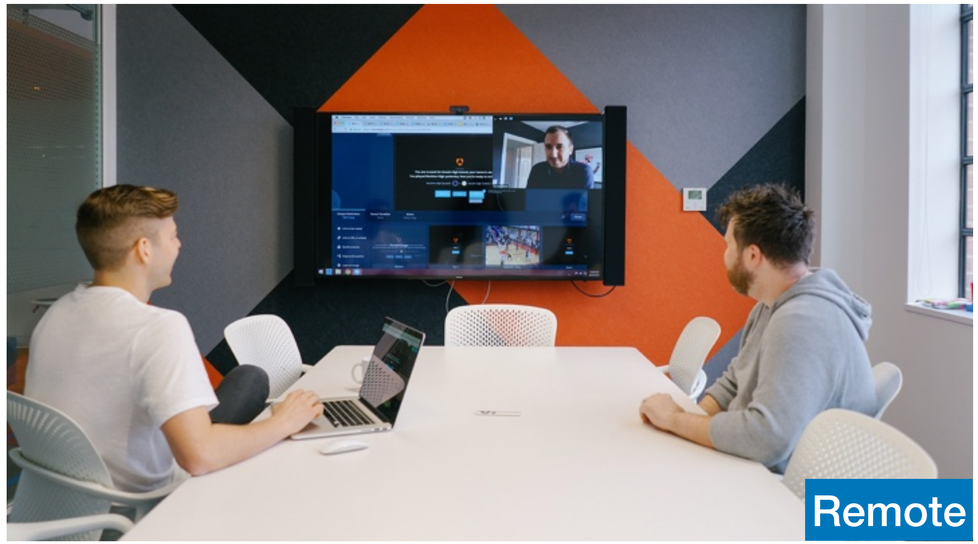
- Limitations:
  - Are subjective
  - Involve several usability experts
  - Cannot find all usability problems
- Thus, **empirical methods** (involving users) are needed
  - Usability tests (engineering approach)
  - Field studies (behavioral approach)
  - Controlled experiments (scientific method)

Empirical Methods

- ! Involving users implies some ethical concerns that require specific cautions
  - ! Asking for explicit consent
  - ! Confidentiality
  - ! Security (avoid any risk)
  - ! Freedom (users may give up at any time)
  - ! Limit stress
- 💡 It’s the system that is under evaluation, not the user!
- Empirical evaluation may be performed in diverse environments ⇒ complementary information
  - In the laboratory (more controlled) vs in the field (more realistic), …
  - In person vs remotely

https://www.nngroup.com/articles/user-research-ethics/

Empirical Methods: environment

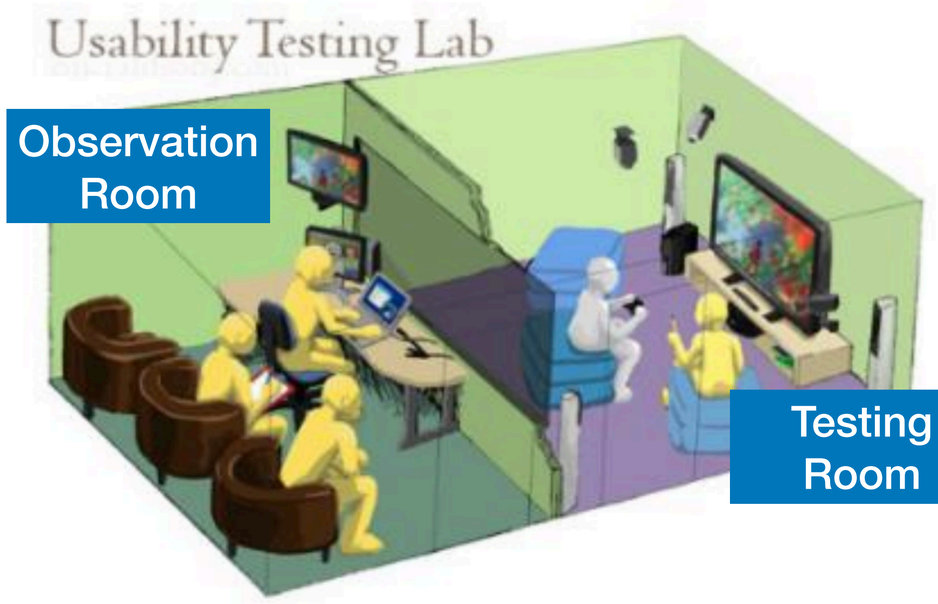


Usability Testing

- **Empirical & (Formative | Summative)**
- Observing real users as they complete tasks
- Can be conducted at all phases of the design (formative or summative)
- Data collected:
  - Quantitative metrics (task success rate, time on task, error rate, etc)
  - Qualitative feedback (user satisfaction with the UI)
- Goal: identify recurring patterns, usability issues, and user sentiments

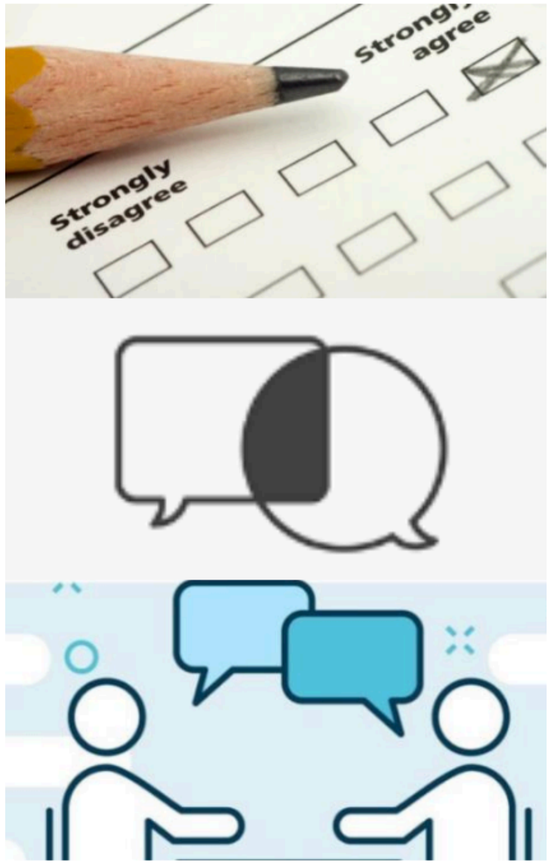
Usability Testing: Observation

- Has many variants, from very simple to very complex (and expensive)
  - Direct / Moderated: A facilitator observes, takes notes and sometimes asks questions
  - Indirect / Unmoderated: Users perform tasks independently, observer follows through audio / video (more complex and time consuming)
  - Think Aloud: Users verbalize their thoughts while performing tasks (unnatural)
  - Logging: users activity is logged by the system
  - …



Usability Testing: Query

- Two main variants to ask users about their experience
  - Survey or Questionnaire
    - Users answer questions post-interaction, often in written form
    - Reach more people, less flexible
  - Interview
    - Researcher asks users about their experience, often guided by a script but free to explore interesting issues in more depth
    - Good balance between richness and replicability
- Data collected: subjective satisfaction, perceived usability, workload ratings



Usability Testing: procedimento

1. **Define test objectives**
   - What will be measured
2. **Define test environment**
   - Where and how the test will take place
3. **Define / recruit target group (test subjects)**
   - Should be representative of the users (number of test subjects, age, gender, skills, …)
4. **Define tasks / script**
   - What will be done by the test subjects
   - Be careful about influencing the results (avoid leading questions, randomise task order, …)
5. **Conduct study / collect data**
   - Video / screen-recording, note taking
   - Think Aloud
   - Questionnaires
6. **Process data**
   - Transcribe data and observations
   - Calculate quantitative metrics and analyse questionnaire results

Usability Questionnaires

- There are several standardised questionnaires available for usability evaluation
  - **Quantification:** They allow usability to be measured objectively
  - **Accuracy:** These questionnaires have been validated — designing a valid and reliable questionnaire is not a trivial task
  - **Generalisation:** Results can be extrapolated beyond the sample group that participated in the study
  - **Communication:** Quantifiable and standardised results are easier to communicate to stakeholders
  - **Comparison:** They make it possible to compare different versions of an interface or competing interfaces

PSSUQ (Post-Study Usability Questionnaire)

- Developed by IBM
- Scoring (lower scores = better usability):
  - **Q1-Q16:** Overall satisfaction with the system
  - **Q1-Q6:** System usefulness
  - **Q7-Q12:** Information quality
  - **Q13-Q15:** Interface quality
- May be affected by acquiescence bias, since all questions are phrased positively

The Post-Study Usability Questionnaire Version 3

			Strongly agree 1	2	3	4	5	6	Strongly disagree 7		NA
1	Overall, I am satisfied with how easy it is to use this system.		o	o	o	o	o	o	o		o
2	It was simple to use this system.		o	o	o	o	o	o	o		o
3	I was able to complete the tasks and scenarios quickly using this system.		o	o	o	o	o	o	o		o
4	I felt comfortable using this system.		o	o	o	o	o	o	o		o
5	It was easy to learn to use this system.		o	o	o	o	o	o	o		o
6	I believe I could become productive quickly using this system.		o	o	o	o	o	o	o		o
7	The system gave error messages that clearly told me how to fix problems.		o	o	o	o	o	o	o		o
8	Whenever I made a mistake using the system, I could recover easily and quickly.		o	o	o	o	o	o	o		o
9	The information (such as online help, on-screen messages and other documentation) provided with this system was clear.		o	o	o	o	o	o	o		o
10	It was easy to find the information I needed.		o	o	o	o	o	o	o		o
11	The information was effective in helping me complete the tasks and scenarios.		o	o	o	o	o	o	o		o
12	The organization of information on the system screens was clear.		o	o	o	o	o	o	o		o
13	The interface* of this system was pleasant.		o	o	o	o	o	o	o		o
14	I liked using the interface of this system.		o	o	o	o	o	o	o		o
15	This system has all the functions and capabilities I expect it to have.		o	o	o	o	o	o	o		o
16	Overall, I am satisfied with this system.		o	o	o	o	o	o	o		o

**The "interface" includes those items that you use to interact with the system. For example, some components of the interface are the keyboard, the mouse, the microphone, and the screens (including their graphics and language).*

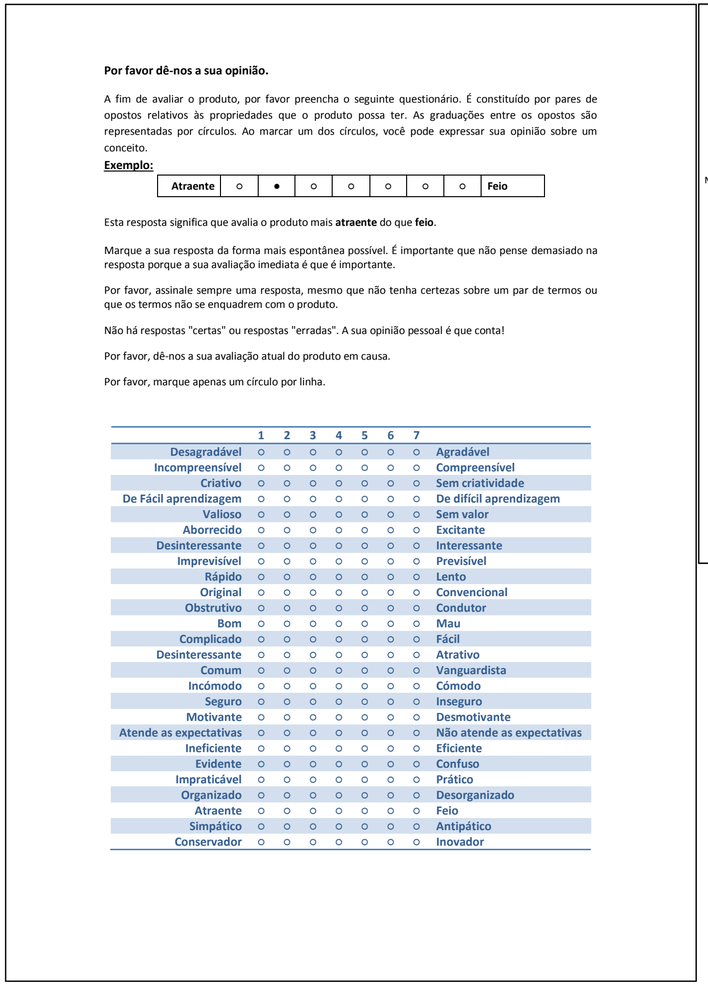
SUS (System Usability Scale)

- One of the most widely used usability questionnaires
- Scoring:
  - **Q1-Q3:** measures overall usability
  - **Q4-Q10:** help assess the perceived ease of learning
  - Adjust and sum all scores
  - Final score between 0 and 100
  - Not a percentage:
    - Average score: ~68; >80: Excellent usability; <51: Poor usability

The System Usability Scale Standard Version			Strongly disagree				Strongly agree
			1	2	3	4	5
1	I think that I would like to use this system.		○	○	○	○	○
2	I found the system unnecessarily complex.		○	○	○	○	○
3	I thought the system was easy to use.		○	○	○	○	○
4	I think that I would need the support of a technical person to be able to use this system.		○	○	○	○	○
5	I found the various functions in the system were well integrated.		○	○	○	○	○
6	I thought there was too much inconsistency in this system.		○	○	○	○	○
7	I would imagine that most people would learn to use this system very quickly.		○	○	○	○	○
8	I found the system very cumbersome to use.		○	○	○	○	○
9	I felt very confident using the system.		○	○	○	○	○
10	I needed to learn a lot of things before I could get going with this system.		○	○	○	○	○

UEQ (User Experience Questionnaire)

- Questionnaire to measure UX
  - Attractiveness
  - Perspicuity
  - Efficiency
  - Dependability
  - Stimulation
  - Novelty
- There is a Portuguese version
- Documentation and tools



Por favor dê-nos a sua opinião.

A fim de avaliar o produto, por favor preencha o seguinte questionário. É constituído por pares de opostos relativos às propriedades que o produto possa ter. As graduações entre os opostos são representadas por círculos. Ao marcar um dos círculos, você pode expressar sua opinião sobre um conceito.

Exemplo:

Atraente	○	●	○	○	○	○	○	Feio

Esta resposta significa que avalia o produto mais **atraente** do que **feio**.

Marque a sua resposta da forma mais espontânea possível. É importante que não pense demasiado na resposta porque a sua avaliação imediata é que é importante.

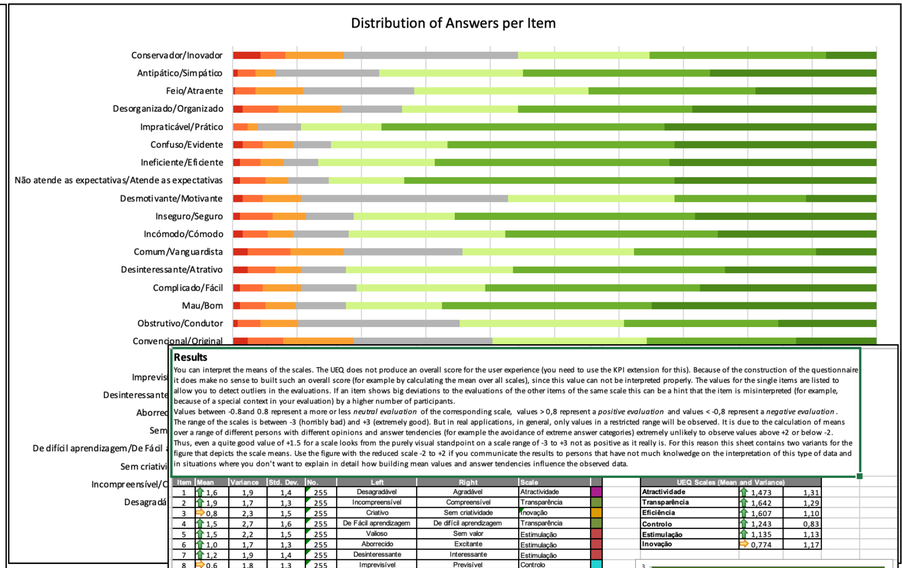
Por favor, assinale sempre uma resposta, mesmo que não tenha certezas sobre um par de termos ou que os termos não se enquadrem com o produto.

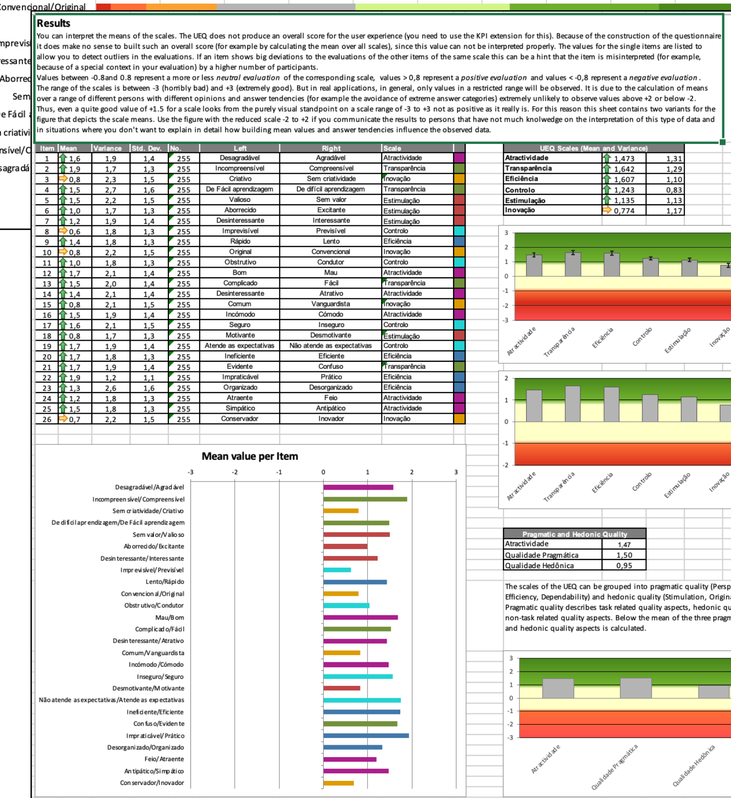
Não há respostas "certas" ou respostas "erradas". A sua opinião pessoal é que conta!

Por favor, dê-nos a sua avaliação atual do produto em causa.

Por favor, marque apenas um círculo por linha.

	1	2	3	4	5	6	7	
Desagradável	○	○	○	○	○	○	○	Agradável
Incompreensível	○	○	○	○	○	○	○	Compreensível
Criativo	○	○	○	○	○	○	○	Sem criatividade
De Fácil aprendizagem	○	○	○	○	○	○	○	De difícil aprendizagem
Valioso	○	○	○	○	○	○	○	Sem valor
Aborrecido	○	○	○	○	○	○	○	Excitante
Desinteressante	○	○	○	○	○	○	○	Interessante
Imprevisível	○	○	○	○	○	○	○	Previsível
Rápido	○	○	○	○	○	○	○	Lento
Original	○	○	○	○	○	○	○	Convencional
Obstrutivo	○	○	○	○	○	○	○	Condutor
Bom	○	○	○	○	○	○	○	Mau
Complicado	○	○	○	○	○	○	○	Fácil
Desinteressante	○	○	○	○	○	○	○	Atrativo
Comum	○	○	○	○	○	○	○	Vanguardista
Incómodo	○	○	○	○	○	○	○	Cómodo
Seguro	○	○	○	○	○	○	○	Inseguro
Motivante	○	○	○	○	○	○	○	Desmotivante
Atende as expectativas	○	○	○	○	○	○	○	Não atende as expectativas
Ineficiente	○	○	○	○	○	○	○	Eficiente
Evidente	○	○	○	○	○	○	○	Confuso
Impraticável	○	○	○	○	○	○	○	Prático
Organizado	○	○	○	○	○	○	○	Desorganizado
Atraente	○	○	○	○	○	○	○	Feio
Simpático	○	○	○	○	○	○	○	Antipático
Conservador	○	○	○	○	○	○	○	Inovador





NASA TLX (Task Load Index)

- Subjective Workload Evaluation
- Based on the weighted average of six sub-scales:
  - Mental Demand
  - Physical Demand
  - Temporal Demand
  - Performance
  - Effort
  - Frustration
- Documentation and tools

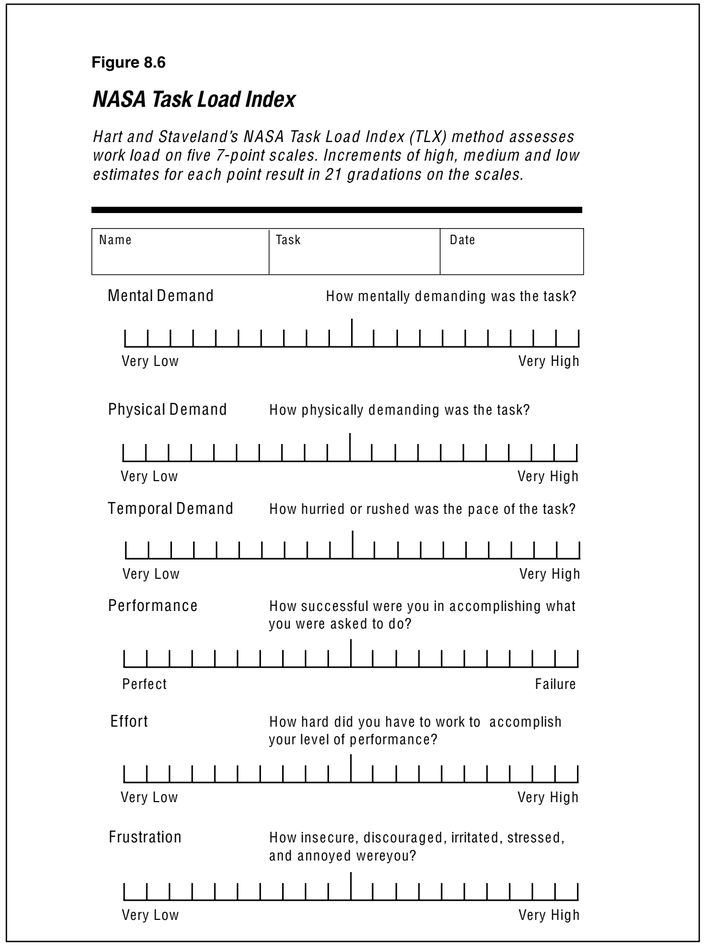


Figure 8.6

NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

Mental Demand — How mentally demanding was the task?
Very Low — Very High

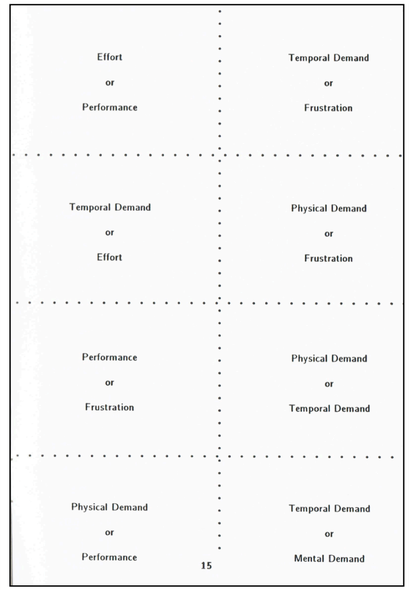
Physical Demand — How physically demanding was the task?
Very Low — Very High

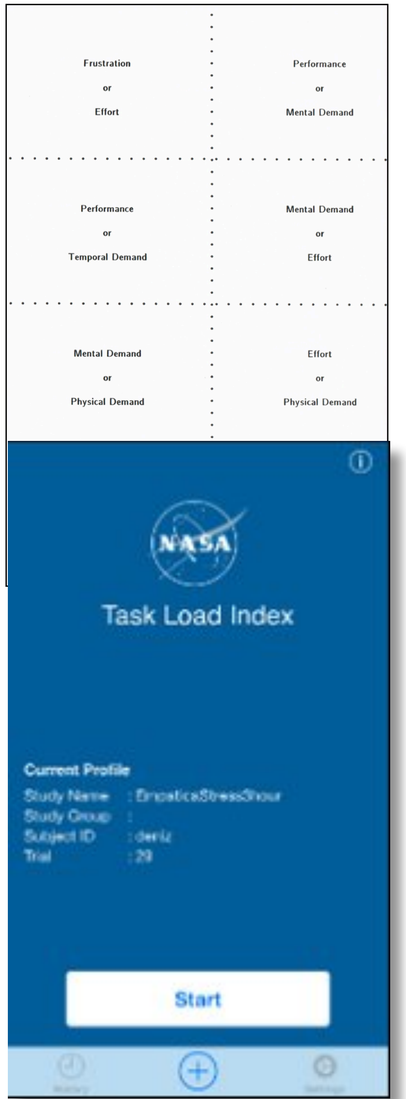
Temporal Demand — How hurried or rushed was the pace of the task?
Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?
Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?
Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?
Very Low — Very High





Field Studies

- **Empirical & Summative**
- **Contextual Inquiry**
  - Observing users in their natural environment while using your product
  - Used for: understanding real-world constraints, workflows, and how the system fits into users' routines
  - Data collected: Observational notes, interviews, usage context insights
- **Diary Studies**
  - Users self-report experiences over time using diaries or logs
  - Used for: longitudinal insights, especially in mobile or wearable tech; capturing evolving experiences
  - Data collected: User-reported entries about usage, problems, satisfaction

Usability Evaluation

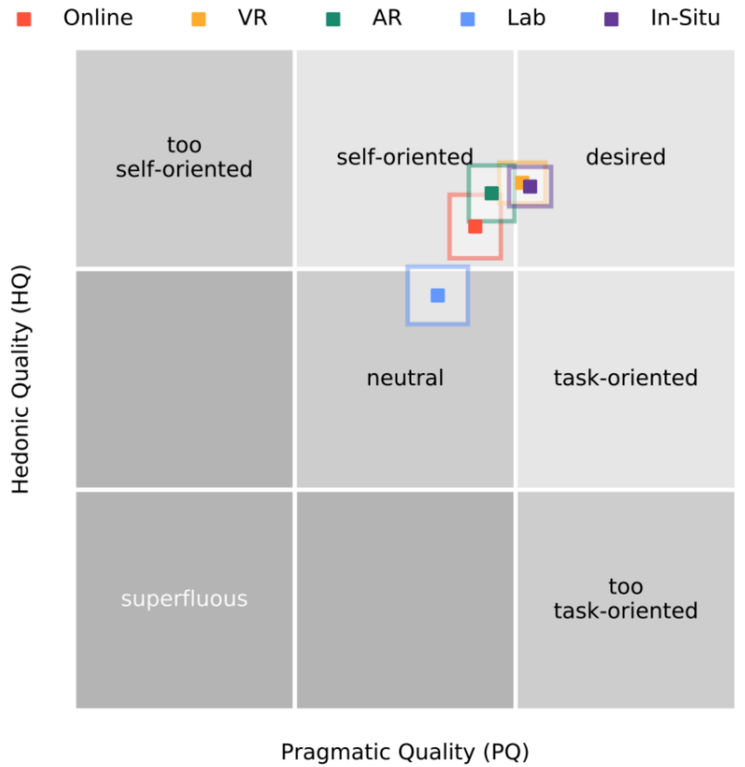
- There are various empirical evaluation methods and environments

? Do they influence the results?

! Yes, participants are not able to ignore the experimental apparatus

! All methods reflect “the truth”, but the influence of external factors determines their validity

💡 Controlled experiments are required to further determine those effects

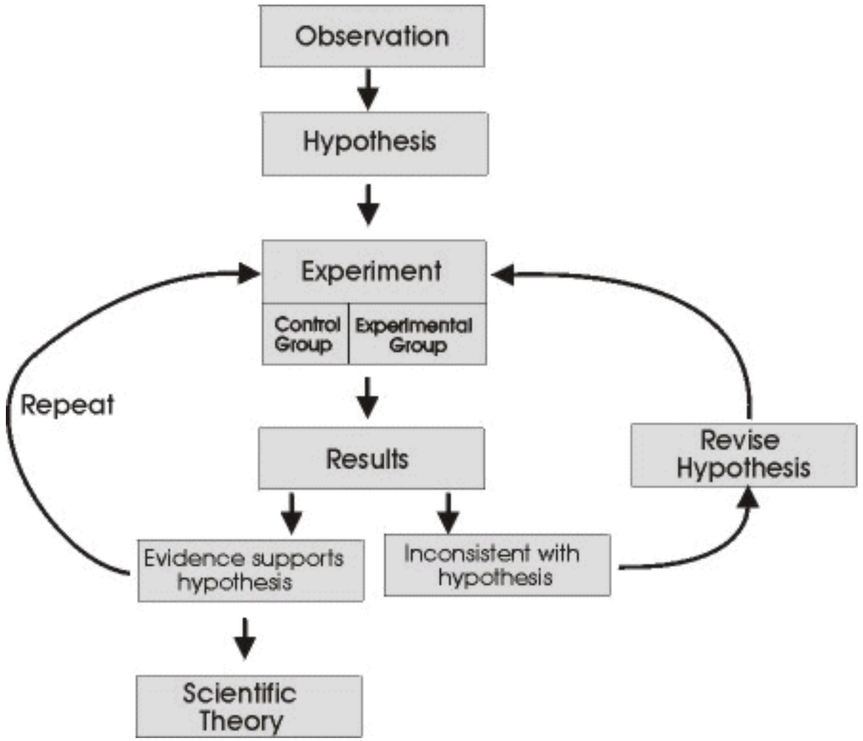


Voit, Alexandra, et al. "Online, vr, ar, lab, and in-situ: Comparison of research methods to evaluate smart artifacts." *Proceedings of the 2019 chi conference on human factors in computing systems*. 2019.

Controlled Experiments

- Probably the only reliable means to isolate cause and effect
- The “work horse” of experimental science, follows the scientific method

1. **Formulate hypotheses**
2. Define input, output and secondary variables
3. Define experimental design (within-groups / between groups)
4. Define protocol
5. Select the participants
6. **Run the experiment** (e.g., a user study or a pilot test)
7. **Analyze the data, repeat**



Summary

- Usability evaluation is essential for identifying problems and improving user experience
- Analytical methods: cost-effective, predictive, and expert-driven
- Empirical methods: user-centered, interpretative, and often more expensive
- Ensures systems meet user needs and expectations
- Helps identify and fix issues early in the development process
- Provides data-driven insights for continuous improvement

👉 Apply these methods in your project!