M'sila University Faculty of Mathematics and Computer

Science Department of Computer Science

HUMAN-MACHINE INTERACTION — EVALUATION OF INTERACTIVE SYSTEMS

Academic Year 2024-2025

Prepared and presented by SAOUDI Lalia

1/ CLASSIFICATION OF EVALUATION METHODS

Evaluation methods are grouped into three main categories:

(1) Methods based on techniques for observing real users and collecting interaction data,

(2) Methods based on expert intervention (in human-machine interaction, cognitive psychology or ergonomics)

(3) Methods referred to as analytical

2. PRINCIPLE OF EVALUATION

- ‘Any evaluation consists of comparing a model of the object being evaluated to a reference model which allows conclusions to be drawn’ definition proposed by [Senach, 1990].

3. Evaluate, what!!!

- **Usability**
  - **the ability to enable the user to easily achieve their goals**
  - **quality of the interface (ergonomics)**
- **Utility**
  - **suitability to the client's high-level objectives**
  - **does the software meet the specifications?**

4. EVALUATE, WHEN

- **During design**
  - iterative design of mock-ups/prototypes
  - user-centred design
  - formative evaluation
- **During implementation**
  - traditional quality control methods
- **During distribution**
  - user satisfaction
  - critical incidents
- **Before a purchase**
  - software comparison
  - summative evaluation

5. EVALUATE THE UTILITY OF A SYSTEM

- **Suitability to the client's high-level objectives**
  - **does the software meet the specifications**
  - **can the user accomplish their task using the system's features**
  - **necessity of clearly formalising the task**

6. EVALUATE THE USABILITY OF A SYSTEM

Usability: ISO 9241

- the degree to which a product can be used, by specified users, to achieve defined goals effectively, efficiently and with satisfaction, in a specified use context

- **Criteria of Usability**
  - effectiveness: achieving the intended outcome
  - efficiency: consumption of a minimum of resources
  - user satisfaction: comfort and subjective evaluation
  - ease of learning
  - ease of appropriation: taking control of the software
  - ease of use
  - reliability: no or few user errors

7. USABILITY EVALUATION TECHNIQUES

- **Experimental evaluation**
  - observation and data collection
  - interviews, questionnaires
- **Analytical evaluation**
  - **Informal approaches:**
    - **expert knowledge:**
      - model of the good interface 'everyone is an expert'
    - **Evaluation grid:**
      - we provide a list of the properties of a good interface, as complete as possible, and we score each of the properties on the list.
      - The evaluator scores on a measurement scale
      - Example from Smith and Mosier 944 rules
    - **Heuristic evaluation approach**
      - Nielsen and Molich 1990 / The criteria of Bastien and Scapin
      - Application of principles or heuristics (guidelines) during design and checks on the prototype.
      - No prior involvement of users in the evaluation
  - **Formal approaches:**
    - methods based on predictive formal models

7. USABILITY EVALUATION TECHNIQUES: 7.1 EMPIRICAL APPROACHES (EXPERIMENTAL EVALUATION)

- **7.1 Experimental evaluation:**
  - The techniques and methods associated with this first class can be used for the evaluation of existing devices when experienced users are available.
  - These approaches are based on techniques observing the real user (end users) and collecting interaction data (questionnaire, interview, verbalisation, eye-tracking, workload estimation, etc.)
  - Among these user-centred approaches, we can mention:
    - **the empirical approaches for usage diagnostics (usable when the HCI is either fully or partially complete),**
    - **approaches focusing on workload estimation and**
    - **approaches based on design testing**

7. USABILITY EVALUATION TECHNIQUES: 7.1.1 EMPIRICAL APPROACHES TO USAGE DIAGNOSIS

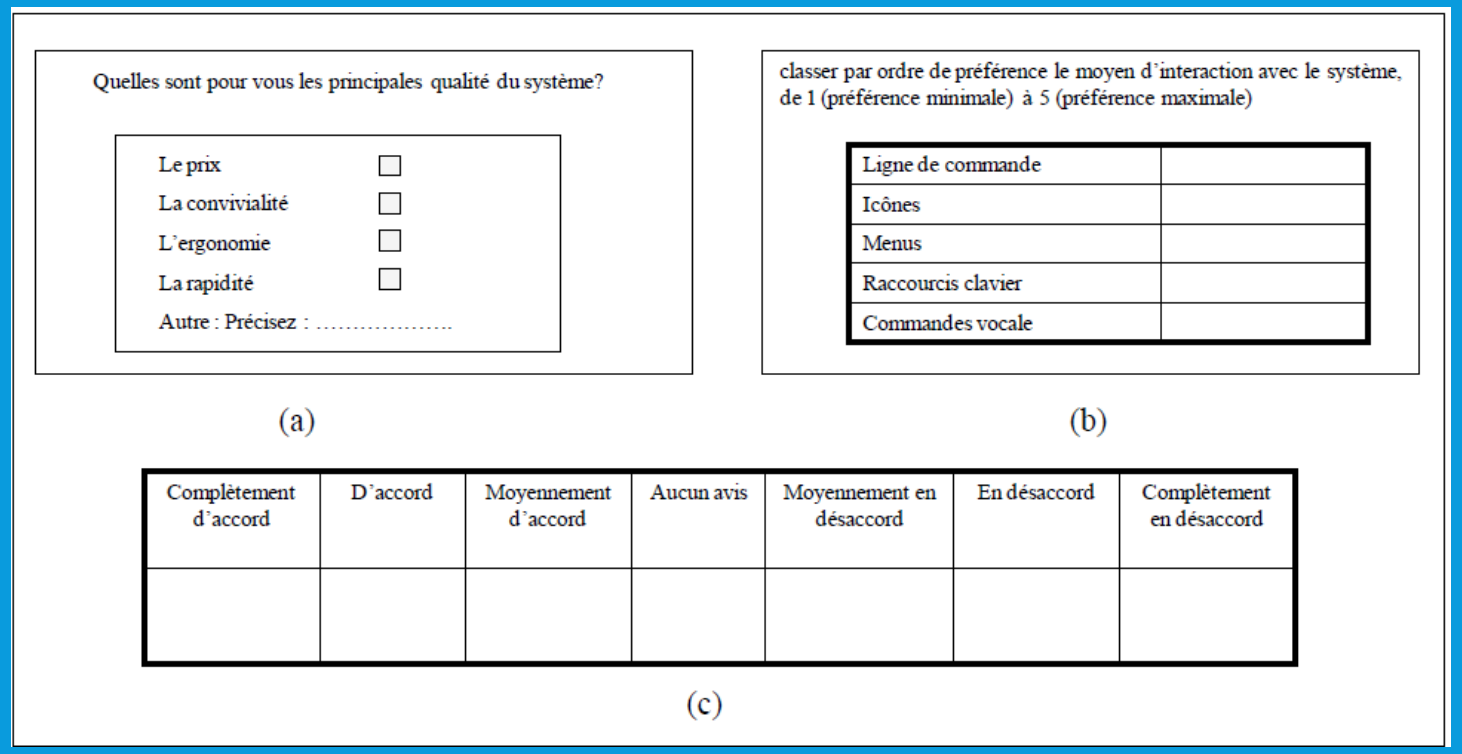
- 7.1.1 EMPIRICAL APPROACHES TO USAGE DIAGNOSIS
  - This approach is only possible if the HCI is operational and ready to be presented to end users
  - This approach is actually very useful when there is prior user experience.
  - The evaluation is essentially based on the collection of information.
    - questionnaires,
    - interviews,
    - the electronic tracker,
    - the eye tracker.

7. USABILITY EVALUATION TECHNIQUES: 7.1.1 EMPIRICAL APPROACHES TO USAGE DIAGNOSIS

- 7.1.1 EMPIRICAL APPROACHES TO USAGE DIAGNOSIS
  - **Usage Questionnaire:**
    - This involves questioning the user about their information needs, inconsistencies in the interface, strengths or weaknesses they felt during their use of the system, and potentially their suggestions on how to correct them [Baccino et al., 2005]. The questionnaire may cover aspects related to the operation of the system and/or aspects related to the ergonomics of the interface.

7. USABILITY EVALUATION TECHNIQUES: 7.1.1 EMPIRICAL APPROACHES TO USAGE DIAGNOSIS

- 7.1.1 EMPIRICAL APPROACHES TO USAGE DIAGNOSIS
  - **Usage Questionnaire:**



Quelles sont pour vous les principales qualité du système?

- Le prix ☐
- La convivialité ☐
- L'ergonomie ☐
- La rapidité ☐
- Autre : Précisez : .....

(a)

classer par ordre de préférence le moyen d'interaction avec le système, de 1 (préférence minimale) à 5 (préférence maximale)

Ligne de commande	
Icônes	
Menus	
Raccourcis clavier	
Commandes vocale	

(b)

Complètement d'accord	D'accord	Moyennement d'accord	Aucun avis	Moyennement en désaccord	En désaccord	Complètement en désaccord

(c)

7. USABILITY EVALUATION TECHNIQUES: 7.1.1 EMPIRICAL APPROACHES TO USAGE DIAGNOSIS

- **7.1.1 Empirical Approaches to Usage Diagnosis**
  - **The interview:**
  - The interview consists of engaging with representative subjects to gain a quick overview of the issues regarding utility and usability encountered by the user during their interaction with the proposed system.
  - It is generally based on techniques used in social psychology [Holyer, 1993].
  - The effectiveness of the interview depends on its preparation and the skills of the evaluator who guides it.
  - The evaluator must prepare and plan in advance the central questions to be asked of the user.

7. USABILITY EVALUATION TECHNIQUES: 7.1.1 EMPIRICAL APPROACHES TO USAGE DIAGNOSIS

- **7.1.1 empirical approaches to usage diagnosis**
  - **The electronic tracker:**
- The electronic tracker or monitoring aims to automatically collect objective data (the actions of users and their repercussions on the system) in real working situations.
- The data collected will be analysed later (manually or automatically)
- The advantage of the electronic tracker is its transparency with respect to the user. Indeed, during the evaluation, the user behaves in a manner that could be described as habitual.
- The disadvantage of electronic trackers lies in the handling of the amount of data it records and in the level of granularity of the analysis [Bastien and Scapin, 2002]

7. USABILITY EVALUATION TECHNIQUES: 7.1.1 EMPIRICAL APPROACHES TO USAGE DIAGNOSIS

- **7.1.1 empirical approaches to usage diagnosis**
  - **The analysis of eye movements**
- The analysis of eye movements or 'eye-tracking' allows the tracking of eye movements and the identification of the gaze location.
- The eye movement technique involves recording the trajectories and fixations generated during reading, information seeking, or inspecting a human-computer interface.
- The advantage of this technique lies in the fact that it allows for a detailed study of how the user searches for and locates useful information depending on various operating situations of the system after data processing.
- One disadvantage of this technique is the limitation of the field of view with most devices.

7. USABILITY EVALUATION TECHNIQUES: 7.1.2 ESTIMATING WORKLOAD

- **7.1.2 Estimating workload**
  - The estimation of workload allows for measuring the level of difficulty associated with the use of the HCI. It provides a qualitative and/or quantitative estimate of the level of activity of the user performing a task.
  - Workload estimation methods can be:
    - **subjective [Cooper and Harper, 1969; Wierwille and Casali, 1983], consisting of asking the user to estimate their workload with subjective scales.**
    - **objective [Sperandio, 1996], utilising measured physical or physiological parameters to estimate workload. Among the objective workload estimation methods, one can mention temporal approaches that define workload as the ratio between the time required for a human operator to perform a task and the time they actually have to execute it.**

7. USABILITY EVALUATION TECHNIQUES: 7.1.2 WORKLOAD ESTIMATION

- **7.1.2 Workload Estimation**
  - The advantage of this method is that it can be used a priori, when the interactive system is at the mock-up or prototyping stage, and a posteriori when it is available,
  - The major disadvantages of these methods are that they only provide an estimate of the workload;
  - for more precise measurements, it involves combining several methods, possibly requiring the participation of specialists and experts in the fields of psychology or physiology

7. USABILITY EVALUATION TECHNIQUES: 7.1.2 WORKLOAD ESTIMATION

- **7.1.3 Design Testing**
  - Design testing methods aim to evaluate and validate an interactive system or human-computer interface through an iterative cycle throughout the development process with representative users if possible.
  - Design tests encompass three complementary approaches:
    - the selection of design alternatives: it is carried out when there are no obvious selection criteria among several possibilities; the collection of empirical data should allow for the prioritisation of the proposed solutions
    - prototyping: it allows for early evaluation, with users, of certain specific aspects of the interface. This technique aims to minimise the development costs of the interface
    - the final testing bench: it is usable once the interface is implemented. It allows for the control of the quality of the final product through a measurement of the overall usability of the system. Objective and subjective data are collected using specially arranged evaluation stations (eye tracker, camera, microphone, etc.).

7. USABILITY EVALUATION TECHNIQUES: 7.2 ANALYTICAL EVALUATION

7.2 EXPERT-BASED APPROACHES

- The evaluation carried out by an ergonomics expert or a human-computer interaction specialist provides a judgment on the ergonomic quality of an interface, comparing the attributes and characteristics of the HCI against ergonomic recommendations.
- we can mention the involvement of a specialist, inspection methods (including the cognitive walkthrough method, guideline recommendations, heuristic evaluation) and evaluation grids.

7. USABILITY EVALUATION TECHNIQUES: 7.2 ANALYTICAL EVALUATION

7.2 .1. THE INVOLVEMENT OF A SPECIALIST

- The aim of this method is to assess the ergonomic quality of the interface by a specialist in human-computer communication and to propose improvements.
- The intervention assessment by a specialist involves conducting an evaluation of the interface and identifying all the potential problems that may be encountered during its use. The involvement of different specialists for the evaluation of an interface shows that each focuses on specific aspects of the interface and bases their assessment on an individual approach [Hammond et al., 1984; Nielsen and Molich, 1990]. Consequently, the ideal would be to involve several specialists (psychologists, cognitive scientists, graphic designers, etc.)

7. USABILITY EVALUATION TECHNIQUES: 7.2 ANALYTICAL EVALUATION

7.2.2. USABILITY INSPECTION METHODS

- Usability inspection methods encompass a range of approaches that rely on the judgement of evaluators, whether they are experts or not in usability.
  - The Cognitive Walkthrough method: this method allows for the evaluation of the system's ease of learning and predicts the problems encountered by the user [Polson et al., 1992; Jacobsen and John, 2000]. By putting oneself in the position of a user exploring the interface while performing predefined tasks,
  - Compliance with recommendations and ergonomic dimensions (ergonomic guides/guidelines): the evaluation of this involves judging the compliance of the interface elements against rules available in various documents, which typically take the form of compilations. Style guides, ergonomic guides [Smith and Mosier, 1986; Bastien and Scapin, 1993; Vanderdonckt, 1994], and standards [ISO, 1996; AFNOR, 2003] can be found.
  - Heuristic evaluation: heuristic evaluation primarily involves conducting a review of the interface by one or more evaluators. By relying on a list of heuristic criteria based on experience in the assessment field focused on usability and taking into account the context of use, the experts (evaluators) compile the issues encountered in relation to these heuristics.

7. USABILITY EVALUATION TECHNIQUES: 7.2 EXPERT-BASED APPROACHES

7.2.3 EVALUATION GRIDS

- Evaluation grids allow for the assessment of the interface according to a set of ergonomic criteria. Using evaluation grids, the evaluator systematically rates the interface on a scale usually consisting of 3 to 5 points, sometimes 10.
- An evaluation grid consists of a set of questions that the evaluator, reviewing the interface, must answer. Typically, the responses to these questions are made via checkboxes or ratings assigned to each ergonomic aspect being studied.
- This method is considered effective since if the evaluator has a set of standard grids [Scapin, 1986; Ravden and Johnson, 1989] along with ergonomic criteria guidelines, [Nielsen, 1993; Vanderdonckt, 1994], they are less likely to overlook important criteria during the evaluation.

7. USABILITY EVALUATION TECHNIQUES: 7.2 ANALYTICAL EVALUATION

7.2.3 EVALUATION GRIDS

	Répondre aux questions ci-après en mettant une croix dans la colonne correspondante					
N°	**Questions**	**Toujours**	**Souvent**	**Quelquefois**	**Jamais**	**Ne sait pas**
Prise en compte de la cohérence						
1	Les couleurs sont-elles utilisées de façon cohérente ?	X				
2	Les abréviations, codes et autres informations sont-ils utilisés de façon cohérente ?		X			
3	La représentation graphique, symboles, icônes et autres informations picturales sont-elles utilisées de façon cohérente ?	X				
4	Le curseur apparaît-il dans la même position initiale pour des affichage similaires					X
.	…					
Prise en compte de la clarté et netteté						
12	Les étapes que peut atteindre l'interface lors de l'exécution d'une tâche sont-elles bien comprises ?	X				
13	L'utilisateur sait-il toujours où il est ?		X			
14	Les opérations dans chaque partie de l'interface sont-elles bien comprises ?			X		
.	…					
	Y a-t-il d'autres commentaires que vous désirez ajouter ?	Bla bla bla…				

7. USABILITY EVALUATION TECHNIQUES: 7.3 ANALYTICAL METHODS

- These approaches, also referred to as model-based approaches [Grislin and Kolski, 1996], are based on formal models of the interface and/or human-computer interaction as well as the gradual implementation of objective metrics. The models are used to predict, based on specifications, certain aspects related to human-computer interaction such as task hierarchy, user action pathways, and the time required to complete a task.
- Two classes of models are distinguished:
  - predictive formal models and
  - interface quality models.

7. USABILITY EVALUATION TECHNIQUES: 7.3 ANALYTICAL METHODS - PREDICTIVE MODELS

Predictive formal models assume that certain user performances can be predicted, and thus considered, during the design and evaluation phase of the interface.

- There are two types of models that can be used as a basis for evaluation:
  - **task models and**
  - **linguistic models.**

7. USABILITY EVALUATION TECHNIQUES: 7.3 ANALYTICAL METHODS - PREDICTIVE MODELS

- **7.3.1 Using task models as a basis for evaluation**
- These models aim to analyse and represent a user's activity in the form of a task structure [Diaper, 1989; Diaper and Stanton, 2004].
- They thus provide a formal basis for the predictive evaluation of performance by offering a measurable description of user behaviour and their performance.
- Among the best known are the following task models:
  - the GOMS model
  - the Keystroke-Level-Model

7. USABILITY EVALUATION TECHNIQUES: 7.3 ANALYTICAL METHODS - PREDICTIVE MODELS

the GOMS model

- **GOMS (Goals, Operators, Methods, and Selector) [Card et al., 1983] describes**
- the cognitive activity of a user (expert) engaged in performing a routine task without error. It is broken down into Goals organised in a hierarchy, Operators corresponding to elementary actions, Methods or procedures for achieving a goal, and Selection rules for methods when multiple solutions are available to achieve the same goal.
- Note that the CPM-GOMS model [John and Kieras, 1996] offers an extension of GOMS in the form of a critical path calculation for performing the task; it describes the user's tasks through a representation in the form of a network.

7. USABILITY EVALUATION TECHNIQUES: 7.3 ANALYTICAL METHODS - PREDICTIVE MODELS

The Keystroke-Level-Model: KLM [Card et al., 1983]:

- essentially allows predicting the execution time of user tasks. A task is analysed at the level of its task-units. For example: entering, moving, deleting… The execution time of each task-unit is estimated by adding the execution times of the primitive operators that compose it.
- This model contains six types of primitive operators, and for each of these operators, an execution time is estimated;
  - K: pressing a key,
  - P: indicating an area on the screen with the mouse,
  - H: directing the hand towards a device,
  - D: drawing a line segment,
  - M: mentally preparing to perform an action,

7. USABILITY EVALUATION TECHNIQUES: 7.3 ANALYTICAL METHODS - PREDICTIVE MODELS

7.3.2 Use of linguistic models as support for evaluation

- These models attempt to produce, on the one hand, an explicit representation of the structure of the interface, and on the other hand, the actions that may be performed by the end users of the HCI, using grammars. These models are usable to describe the interaction tasks between the user and the interface, primarily for the evaluation of the command language and its coherence.