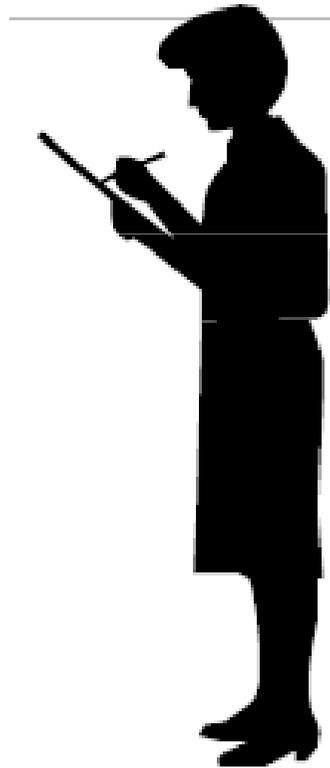


Chapitre VII

Méthodologies d'Evaluation des IHMs



Plan du cours

- I. Introduction
- II. Quand évaluer ?
- III. Variables cibles
- IV. Modalités d'évaluation
 - 1. Approches analytiques
 - 2. Approches empiriques
- V. Préparation des utilisateurs aux tests
- VI. Critères de qualités de Schneiderman
- VII. Norme ISO 9241-11
- VIII. Conclusion

I. Introduction

L'évaluation est une phase primordiale dans la boucle de conception des IHM.

- ✓ Elle vise à **identifier** et à **prévoir** les difficultés que rencontrent les utilisateurs et à en caractériser les **points forts** et les **points faibles**.
- ✓ Elle permet de **vérifier** et **valider** un système interactif.

Un système est dit **vérifié** s'il correspond aux **spécifications** issues de **l'analyse des besoins**.

Un système est dit **validé** s'il correspond aux **besoins** tout en respectant les **contraintes du domaine d'application**.

L'évaluation vérifie donc **l'utilisabilité** et **l'utilité** de l'IHM.

II. Quand Evaluer ?

1. **En cours de conception** : test papier et surtout maquettes pour décider du design à adopter.
2. **En cours de spécification (réalisation)**
 - ❖ Expérimentation par prototypage (évaluation de fonctionnement, vérification de comportement, et de performances à divers niveaux),
 - ❖ Prototypes jetables ou évolution vers le produit fini.
3. **Avant diffusion puis après**: tests, enquêtes, expertises chez les utilisateurs, boites de suggestions en lignes, etc.

III.

variables cibles

mesures de performances

- taux d'erreur
- durée d'exécution d'une tâche
- durée de lecture documentation
- demandes d'aide
- fonctions inutilisées
- durée d'apprentissage

variables subjectives

- confort d'usage
- suggestions préférences
- esthétique
- facilité d'apprentissage
- décision d'achat

IV. Modalités d'évaluation

Contrôle de qualité avec un modèle défini a priori

Expérimentations menées avec des utilisateurs

APPROCHES ANALYTIQUES

APPROCHES EMPIRIQUES

FORMELLES
à caractère théorique

Approches **INFORMELLES**

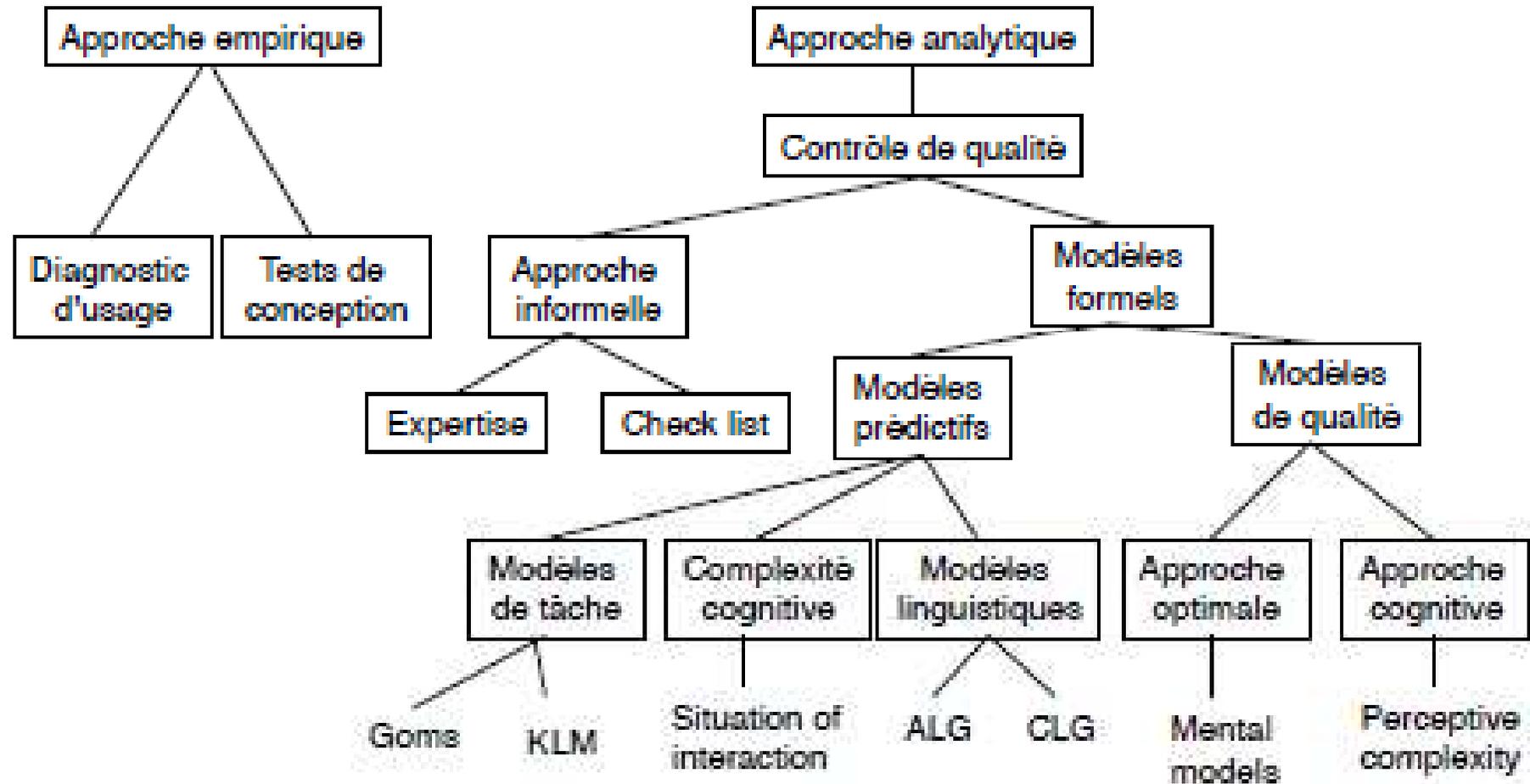
Diagnostic d'usage

Tests de conception



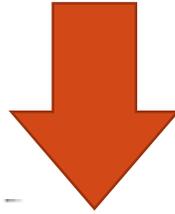
Classification des méthodes d'évaluation

(Senach, 90)

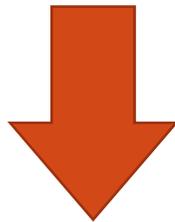


B. Senach, "Evaluation ergonomique des interfaces homme-machine une revue de la littérature", Rapport de recherche INRIA n° 1180, Programme 8, Communication Homme-Machine, Mars 1990.

APPROCHES ANALYTIQUES



Approches INFORMELLES



1. Expertise
2. Grille d'évaluation
3. Evaluation heuristique

Approches Analytique

1. Approches informelles

1. **Expertise** « le savoir de l'expert » : modèle de la bonne interface.



2. **Grille d'évaluation**

On fournit une liste des propriétés d'une bonne interface, aussi complète que possible, et on note chacune des propriétés de la liste.

L'évaluateur **note sur une échelle** de mesure.

Grille d'évaluation

	Prise en compte de l'expérience de l'utilisateur	1	2	3	4	5
1	L'information est-elle présentée dans le même état d'esprit que l'utilisateur s'en fait ?			X		
2	L'affichage graphique est-il compatible avec la représentation que se fait l'utilisateur du système ?	X				
3	La structure du système est-elle adaptée à la perception de la tâche de l'utilisateur ?				X	
4	Le déroulement des activités exigées pour effectuer une tâche est-il en accord avec les attentes de l'utilisateur ?		X			
5	Le système travaille-t-il dans la même voie que la pensée de l'utilisateur ?					X
6	Estimation globale du système en terme de prise en compte de l'expérience de l'utilisateur ?					X
	Clarté visuelle					
1	Chaque écran est-il clairement identifié (titre informatif ou description) ?				X	
2	L'organisation des informations sur l'écran est-elle toujours logique ?			X		
3	...					

Approches Analytique

1. Approches informelles

3. Évaluation Heuristique

- Nielsen et Molich (1993) ont dégagé **10 heuristiques** qui couvrent en pratique tous les problèmes (chapitre 3).
- Bastien & Scapin (1993).



Procédure

1. Plusieurs **experts** en évaluation **analysent indépendamment** l'IHM en utilisant les heuristiques.
2. Les **problèmes** identifiées sont **rassemblées** dans une **liste** unique puis discutées en groupe.

Nielsen's Heuristic Evaluation

Inspiré de: Nielsen, J. (1993) *Usability Engineering*. Academic Press. Chapter 5, p. 115. | [Au sujet du questionnaire quest.cgi](#)

Please evaluate le système according to Nielsen's usability heuristics.

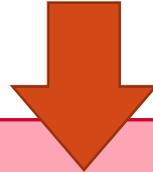
		1	2	3	4	5	6	7	NA
1. Simple and Natural Dialogue <input type="checkbox"/>	Mauvais	<input type="radio"/>	Bien <input type="radio"/>						
2. Speak the Users' Language <input type="checkbox"/>	Mauvais	<input type="radio"/>	Bien <input type="radio"/>						
3. Minimize User Memory Load <input type="checkbox"/>	Mauvais	<input type="radio"/>	Bien <input type="radio"/>						
4. Consistency <input type="checkbox"/>	Mauvais	<input type="radio"/>	Bien <input type="radio"/>						
5. Feedback <input type="checkbox"/>	Mauvais	<input type="radio"/>	Bien <input type="radio"/>						
6. Clearly Marked Exits <input type="checkbox"/>	Mauvais	<input type="radio"/>	Bien <input type="radio"/>						
7. Shortcuts <input type="checkbox"/>	Mauvais	<input type="radio"/>	Bien <input type="radio"/>						
8. Good Error Messages <input type="checkbox"/>	Mauvais	<input type="radio"/>	Bien <input type="radio"/>						
9. Prevent Errors <input type="checkbox"/>	Mauvais	<input type="radio"/>	Bien <input type="radio"/>						
10. Help and Documentation <input type="checkbox"/>	Mauvais	<input type="radio"/>	Bien <input type="radio"/>						
		1	2	3	4	5	6	7	NA

[Haut](#) | [Retour au site](#)

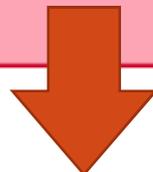
APPROCHES ANALYTIQUES



FORMELLES
à caractère théorique



**MODÈLES
PRÉDICTIFS
DES
PERFORMANCES**



analyse des tâches
approche linguistique
complexité cognitive

Modèles prédictifs des performances

a. Analyse des tâches

Description logique des activités à réaliser pour atteindre les objectifs des utilisateurs lors d'une interaction HM.

- **L'analyse** du comportement rationnel de l'utilisateur en utilisant la méthode **GOMS** (Goal, Operator, Method, Selection) et
- **l'évaluation** des actions physiques de l'utilisateur en utilisant la méthode de **keystroke**;

1. **Buts** (Goal): définit un état recherché,
2. **Méthodes** (Method): décrivent un procédé permettant d'atteindre un but,
3. **Opérateurs** (Operator): actions élémentaires provoquant un changement d'état,
4. **Règles de sélection** (Selection rules) pour le choix entre méthodes.

La méthode Keystroke

Permet une **approximation** du **temps** que prend une opération (tâche). Elle prend en compte les actions physiques que les utilisateurs exécutent pour mettre en œuvre une commande.

Le modèle de cette technique est basé sur le fait que chaque type d'opération prend un temps différent:

K = 0,2 s : **Keying** (taper une touche au clavier)

P = 1,1 s : **Pointing** (déplacer la souris vers une cible)

H = 0,4 s : **Homing** (aller de la souris au clavier vice et versa)

M = 1,35 s : **Processus mental** pour préparer la prochaine étape

Le temps d'exécution (T) d'une tâche est la **somme** des temps requis pour chaque opération.

Ces chiffres varient selon l'habileté de l'utilisateur, mais c'est une moyenne représentative de la réalité.

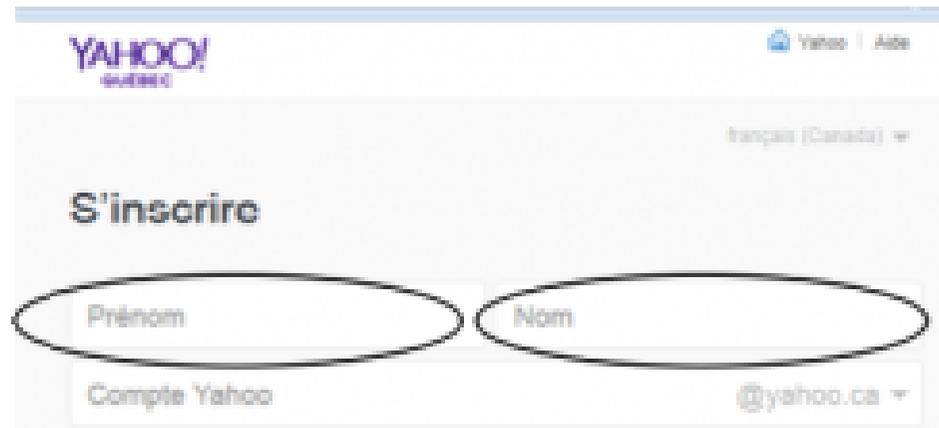
De plus, entrer une séquence de caractères aléatoires prend plus de temps que de taper du texte.

Exemple

But : Remplir les cases Prénom et le Nom dans ce formulaire d'inscription.

Méthode :

- Prendre la souris (H)
- Sélectionner la case « prénom » (P)
- Se souvenir de son prénom (M)
- Écrire les lettres (K)
- Reprendre la souris (H)
- Sélectionner la case « nom » (P)
- Se souvenir de son nom (M)
- Écrire les lettres (K)
- Retour à la fin de la tâche (H).



Opérateurs : déplacement du curseur, clic de souris, enfoncement d'une touche du clavier, sélection avec la souris.

Sélection : règle : entrée de données.

$$T = H + M + P + K + H + M + 5K + H + P + K + H + M + 3K + H + M + P + K$$

$$T = 5H + 4M + 3P + 11K$$

$$T = 5 * 0.4 + 4 * 1,35 + 3 * 1.1 + 11 * 0,2$$

$$T = 2 + 5,4 + 3,3 + 2.2 = 12,9 \text{ sec.}$$

Modèles prédictifs des performances

b. Modèle Linguistique (Action Language Grammar: ALG (Reisner))

C'est la présentation des **actions** de l'utilisateur sous forme de **grammaire**.

- Un ensemble de règles de production de type :
POUR effectuer telle action **FAIRE** telles opérations.
- Définit les règles d'utilisation d'où on tire 3 indices pour caractériser l'interface :
 - ❖ La **complexité du langage** (nombre d'actions pour atteindre un but),
 - ❖ La **simplicité des procédures** (longueur des séquences pour une tâche donnée),
 - ❖ La **cohérence des structures** (nombre de règles non nécessaires et nombre de règles pour des séquences terminales similaires).

Modèles prédictifs des performances

c. Complexité Cognitive

Permet de comparer les mérites relatifs de plusieurs conceptions en matière de **transfert** de **connaissances** et de **difficulté d'apprentissage**.

- Le modèle du **processus mental** s'appuie sur un système de production de règles de la forme : IF (condition) THEN (action),
- Le modèle de la **tâche**, inspiré de GOMS, utilise une décomposition hiérarchique des tâches.

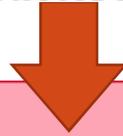
La **complexité de l'interface** est donc évaluée en terme de:

- ❖ Nombre total de règles de production pour modéliser la tâche,
- ❖ Nombre de productions déclenchées,
- ❖ Nombre maximal de buts en mémoire pour une fonction donnée ainsi que la pile de buts maintenus en mémoire de travail.

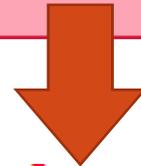
APPROCHES ANALYTIQUES



FORMELLES
à caractère théorique



**MODÈLES
DE
QUALITÉ**



approche cognitive

- modèles mentaux
- cohérence interne et externe

approche perceptive

- complexité perceptive
- qualité d'affichage

Modèles de Qualité

1. Approche Cognitive

cette approche fait appel aux **modèles mentaux** où d'une manière générale certaines conditions doivent être remplies telles que :

- la cohérence interne (il n'existe pas de contradiction à l'intérieur du modèle),
- la validité (le modèle correspond à la réalité de ce qu'il représente), et
- l'intégration (le modèle s'accorde avec les autres connaissances).

Modèles de Qualité

2. Approche Perceptive

il s'agit de prendre en considération la **complexité perceptive** qui concerne principalement la qualité d'affichage ainsi que d'autres critères ergonomiques tels que les groupements, la clarté, la charge mentale, l'homogénéité et la consistance.

APPROCHES EMPIRIQUES



Diagnostic d'usage



- a. Questionnaires
- b. Mouchards électroniques
- c. La méthode « Think aloud »
- d. Evaluation Coopérative

Approches Empiriques

1. Diagnostic d'usage

a. Questionnaires

- ✓ Vérifier la satisfaction des besoins utilisateurs,
- ✓ Appréciations subjectives possibles (Q. ouvertes),



Inconvénients

- Difficultés des choix et de l'élaboration des questions,
- Difficultés d'analyser les réponses pour dégager l'ensemble des corrections à établir dans le système interactif.

Exemples de questionnaires sur ce site

<http://oldwww.acm.org/perlman/question.cgi?form=CSUQ>

Veillez évaluer la facilité d'utilisation.

	1	2	3	4	5	6	7	NA
1. En général, je suis satisfait(e) de la facilité d'utilisation de ce système ☑	Désaccord	<input type="radio"/>	Accord <input type="radio"/>					
2. Ce système est simple à utiliser ☑	Désaccord	<input type="radio"/>	Accord <input type="radio"/>					
3. J'ai complété mon travail correctement en utilisant ce système ☑	Désaccord	<input type="radio"/>	Accord <input type="radio"/>					
4. J'ai été en mesure de compléter rapidement ma tâche avec ce système ☑	Désaccord	<input type="radio"/>	Accord <input type="radio"/>					
5. J'ai complété mon travail efficacement en utilisant ce système ☑	Désaccord	<input type="radio"/>	Accord <input type="radio"/>					
6. Je me sens à l'aise avec ce système ☑	Désaccord	<input type="radio"/>	Accord <input type="radio"/>					
7. J'ai eu de la facilité à apprendre comment utiliser ce système ☑	Désaccord	<input type="radio"/>	Accord <input type="radio"/>					
8. Je crois être devenu(e) rapidement efficace en utilisant ce système ☑	Désaccord	<input type="radio"/>	Accord <input type="radio"/>					
9. Les messages d'erreur présentés par ce système m'indiquent clairement comment résoudre les problèmes ☑	Désaccord	<input type="radio"/>	Accord <input type="radio"/>					
10. Lorsque j'ai fait une erreur d'utilisation sur ce système, il m'a été facile et rapide de la corriger ☑	Désaccord	<input type="radio"/>	Accord <input type="radio"/>					
11. Les outils d'aide disponibles sur ce système (tels que l'aide en ligne, les messages à l'écran et autres informations) sont utiles ☑	Désaccord	<input type="radio"/>	Accord <input type="radio"/>					
12. J'ai facilement trouvé l'information que je cherchais ☑	Désaccord	<input type="radio"/>	Accord <input type="radio"/>					
13. L'information fournie avec ce système est facile à comprendre ☑	Désaccord	<input type="radio"/>	Accord <input type="radio"/>					
14. L'information disponible sur ce système contribue à me soutenir dans la réalisation des tâches et des scénarios ☑	Désaccord	<input type="radio"/>	Accord <input type="radio"/>					
15. L'organisation de l'information dans les écrans du système est claire ☑	Désaccord	<input type="radio"/>	Accord <input type="radio"/>					
16. L'interface de ce système est agréable ☑	Désaccord	<input type="radio"/>	Accord <input type="radio"/>					
17. J'aime utiliser l'interface de ce système ☑	Désaccord	<input type="radio"/>	Accord <input type="radio"/>					
18. Ce système possède toutes les fonctions et le potentiel correspondant à mes attentes ☑	Désaccord	<input type="radio"/>	Accord <input type="radio"/>					
19. En général, je suis satisfait(e) de ce système ☑	Désaccord	<input type="radio"/>	Accord <input type="radio"/>					

Approches Empiriques

1. Diagnostic d'usage

b. Mouchards électroniques

- Recueil automatique, en situation réelle, des actions de l'utilisateur et leur effet sur le système lors de l'interaction avec le système évalué,
- Analyse de données enregistrées.



Inconvénients

- Quantité importante de données enregistrées pour l'évaluation
- Complexité de la mise en place pour le processus d'évaluation.

Approches Empiriques

1. Diagnostic d'usage

C. La méthode « Think aloud »

- Méthode d'évaluation très simple et puissante,
- Utilisable aussi pour tester la documentation fournie,
- Le concepteur se place derrière l'utilisateur et encourage celui-ci à penser à haute voix pendant qu'il utilise le système.



Quelques règles simples

- ❖ L'utilisateur aura été mis en confiance : (ce n'est pas l'utilisateur qu'on teste, c'est le système) et les conditions du test lui sont expliquées et justifiées,
- ❖ Le concepteur ne donne pas d'aides à l'utilisateur (même en cas de demandes) mais note toutes les demandes et interrogations,
- ❖ Le concepteur n'intervient qu'en cas d'impasse grave (problème à résoudre).

Approches Empiriques

1. Diagnostic d'usage

d. Evaluation Coopérative

1. Recruter des utilisateurs/essayers

- ❖ définir le public cible (futurs utilisateurs ou profils types),
- ❖ choisir la cible des tests.

2. Préparer et choisir les activités de test (spécifiques mais représentatives)

- ❖ déterminer la durée d'activité,
- ❖ préparer un descriptif d'activité.

3. Organiser la session de test

- ❖ Enregistrer les réactions des utilisateurs (vidéo, audio, fichier log, ...),
- ❖ Laisser parler, agir les utilisateurs tests et les questionner.

4. Analyser les résultats

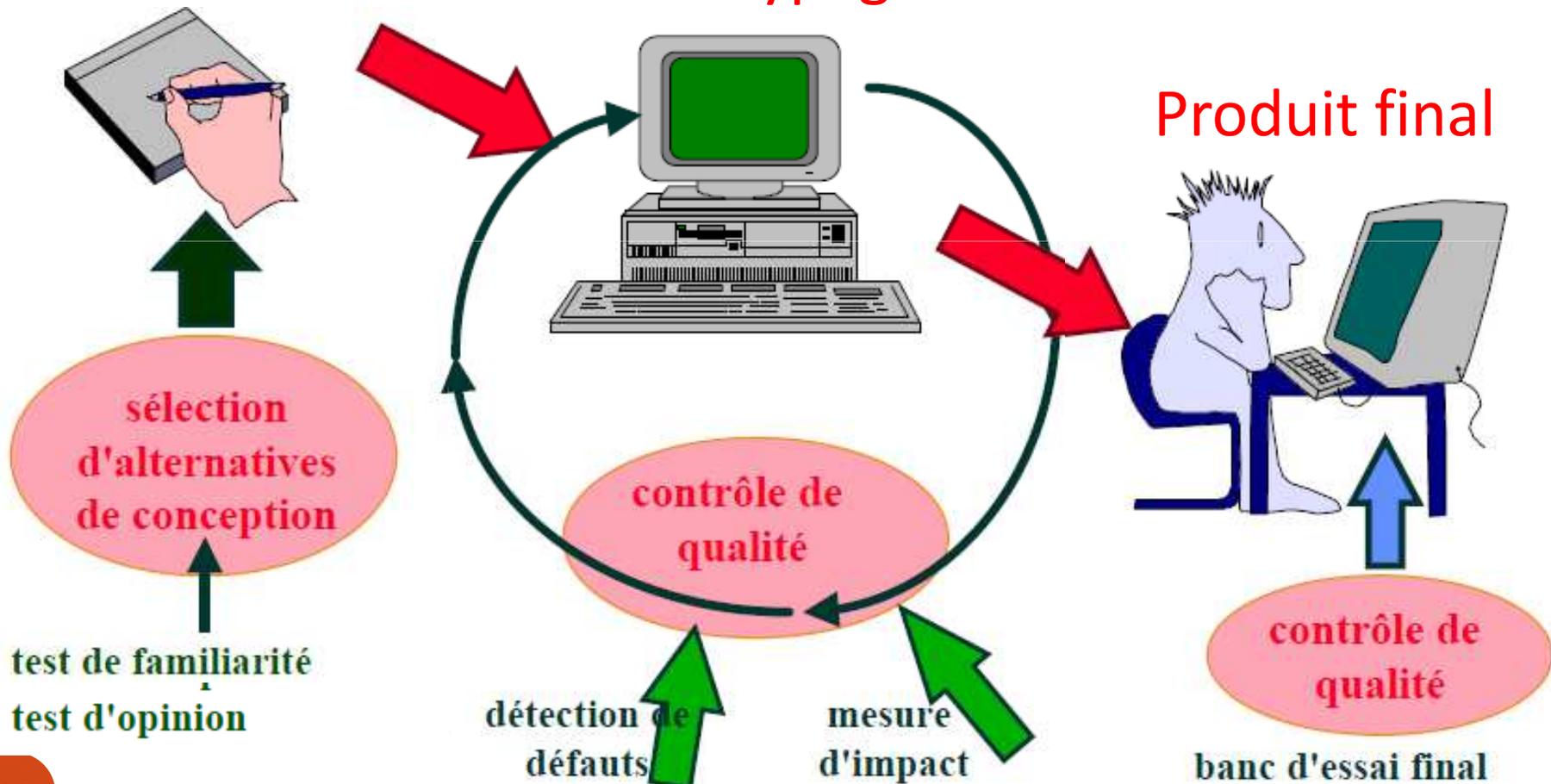
Approches Empiriques

2. Tests de Conception (Senach, 1990)

Maquettage

Prototypage

Produit final



V. Préparation des utilisateurs aux tests

Il faut préciser clairement aux testeurs:

- Évaluation du système et non de l'utilisateur;
- Le système n'est pas développé par le testeur;
- L'étude est confidentielle;
- La participation de l'utilisateur est volontaire;
- Les résultats sont anonymes;
- Vérifier la compréhension des consignes;

Pendant le test

- Ne pas porter de jugement;
- Ne pas l'aider;

Après le test

Debriefing (réunion) avec l'utilisateur.

Un test utilisateur peut s'effectuer sous différentes formes et avoir lieu :

- Dans un laboratoire d'utilisabilité,
- Dans les bureaux des concepteurs,
- Sur la place de travail de l'utilisateur,
- Dans un endroit neutre,
- etc.



Le choix dépend du **contexte** spécifique du système à évaluer, des **informations** que l'on souhaite recueillir (objectifs de l'évaluation) et de la **phase du cycle** de développement dans laquelle on se trouve (stade initial, intermédiaire, ou proche de l'état final).

VI. Critères de Qualité de Shneiderman

Shneiderman propose cinq critères pour évaluer les caractéristiques de la qualité d'une interface:



1. Temps d'apprentissage des dispositifs nécessaires à l'exécution d'une tâche.
2. Rapidité d'exécution d'une tâche.
3. Taux d'erreurs effectuées par l'utilisateur. On évaluera :
 - ✓ la fréquence des erreurs. On distingue les erreurs d'exécution liées à des défauts de manipulation et les erreurs d'intention lorsque l'utilisateur sélectionne une commande inappropriée,
 - ✓ le temps de correction (diagnostic, recommandation).

4. **Période de rémanence** durant laquelle un utilisateur conserve la connaissance acquise (heures, jours, semaines, ...). C'est une fonction :
 - ✓ directe de **l'effort cognitif** requis pour planifier une séquence d'actions qui réalisent une intention donnée ,
 - ✓ indirecte du **temps d'apprentissage** et de la **fréquence d'utilisation**.

5. **Satisfaction subjective** à utiliser le système. Elle peut se traduire par un sentiment de confort et d'enrichissement.

6. **Degré de couverture** (n'est pas cité par Shneiderman) des dispositifs de l'interface par rapport aux actions qui font partie de la tâche de l'utilisateur.

Les **critères de Schneiderman** doivent être **pondérés** suivant les cas:

- Pour un **distributeur de billet**: taux d'erreur et temps d'apprentissage sont importants.
- Pour un **jeu vidéo**: temps d'apprentissage et satisfaction sont importants.
- Pour un **éditeur de code source**: rapidité d'exécution est importante.

VII. Norme ISO 9241-11 recommande



Mesurer l'utilisabilité

1. Efficacité

Vérifier que les **objectifs** visés par l'utilisateur sont **atteints**.

2. Efficience

Déterminer les **ressources** nécessaires pour atteindre les objectifs.

ex. le temps mis par l'utilisateur pour réaliser une tâche.

3. Satisfaction

Déterminer si le système est **agréable** à utiliser.

ex. en décomptant le nombre de remarques négatives émises par les utilisateurs lors du test.

Questions ...