



# Interaction Homme-Machine

Jean-Yves Antoine

<http://www.info.univ-tours.fr/~antoine/>

# Interaction Homme-Machine

## Chapitre 7 – Évaluation des systèmes interactifs

*Le cycle logiciel concerne également l'interaction avec l'utilisateur*

# EVALUATION - Objectifs

## 7.1. Notions

- 7.1.1. Intégration d'une stratégie d'évaluation dans le cycle de vie logiciel
- 7.1.2. Connaître les différentes techniques d'évaluation ergonomique
- 7.1.3. Connaître les règles déontologiques de base en matière d'expériences avec utilisateurs humains
- 7.1.4. Connaître les modèles prédictifs (au moins en théorie) d'estimation de temps d'exécution d'une tâche

## 7.2. Pratiques

- 7.2.1. Savoir définir et conduire un plan d'évaluation en réponse à un besoin donné
- 7.2.2. Savoir mettre en œuvre deux techniques particulières d'évaluation heuristique (promenade cognitive) ou avec test utilisateur (*think aloud*)

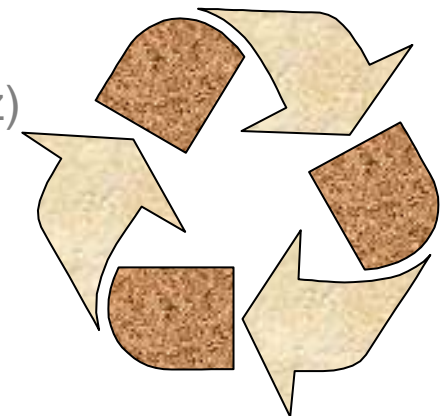
# EVALUATION DE L'INTERACTION

## Pourquoi évaluer ?

- L'intuition du concepteur du système ne peut suffire
- Modélisation formelle du système et de l'interaction ne couvre pas tous les choix de conception
- Les recommandations (*guidelines*) restent des garde-fous et des bonnes pratiques trop générales pour couvrir tous les aspects d'une interaction spécifique.

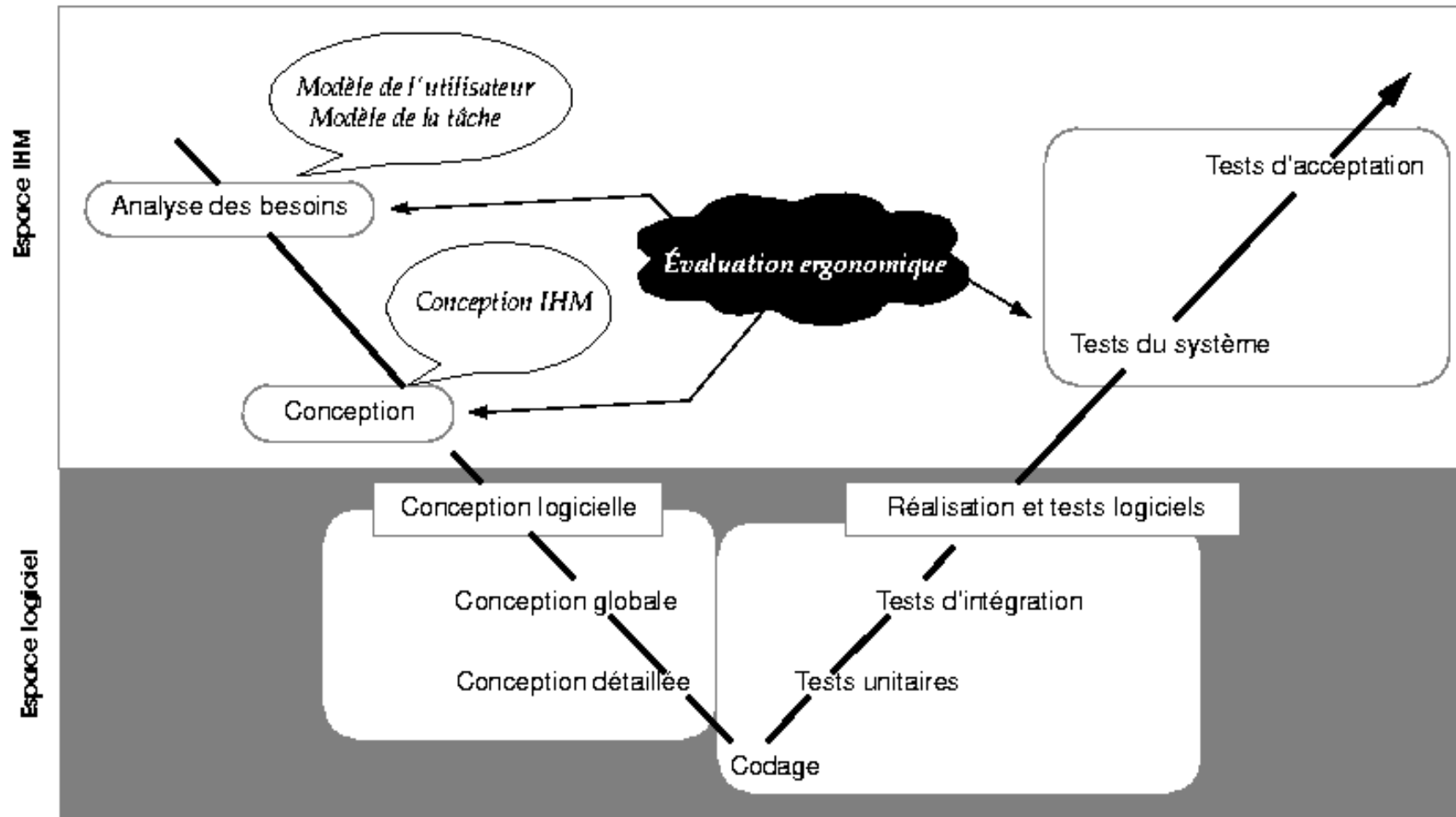
## Quand évaluer ? Le cycle de vie logiciel concerne aussi l'interaction

- Cycle en spirale avec prototypage
- **Evaluation à toutes les étapes** du développement !!!!
- **Prototype** papier vs. Prototype interactif (système ou WOz)
  - Prototype papier se focalise sur l'essentiel
  - Prototype papier plus artificiel (basse fidélité)



# EVALUATION DE L'INTERACTION

## Quand évaluer ?



D'après L. Nigay (ESSI3-IHM)

# EVALUATION DE L'INTERACTION

## Une évaluation ne donne de résultat que si elle est bien préparée

- Quelques questions à se poser...

[Basili *et al*, 1994]

- Quels sont les buts généraux de l'évaluation ?
- Quelles sont les questions spécifiques pour lesquelles on veut obtenir une réponse grâce à l'évaluation ?
- Quel est le paradigme et les techniques de tests nécessaires pour atteindre ces objectifs ?
- Comment organiser en pratique l'évaluation : recrutement des utilisateurs, préparation des utilisateurs, dispositifs de recueil...
- S'assurer du respect des règles déontologiques en vigueur
- Comment dépouiller, interpréter et présenter les données recueillies ?

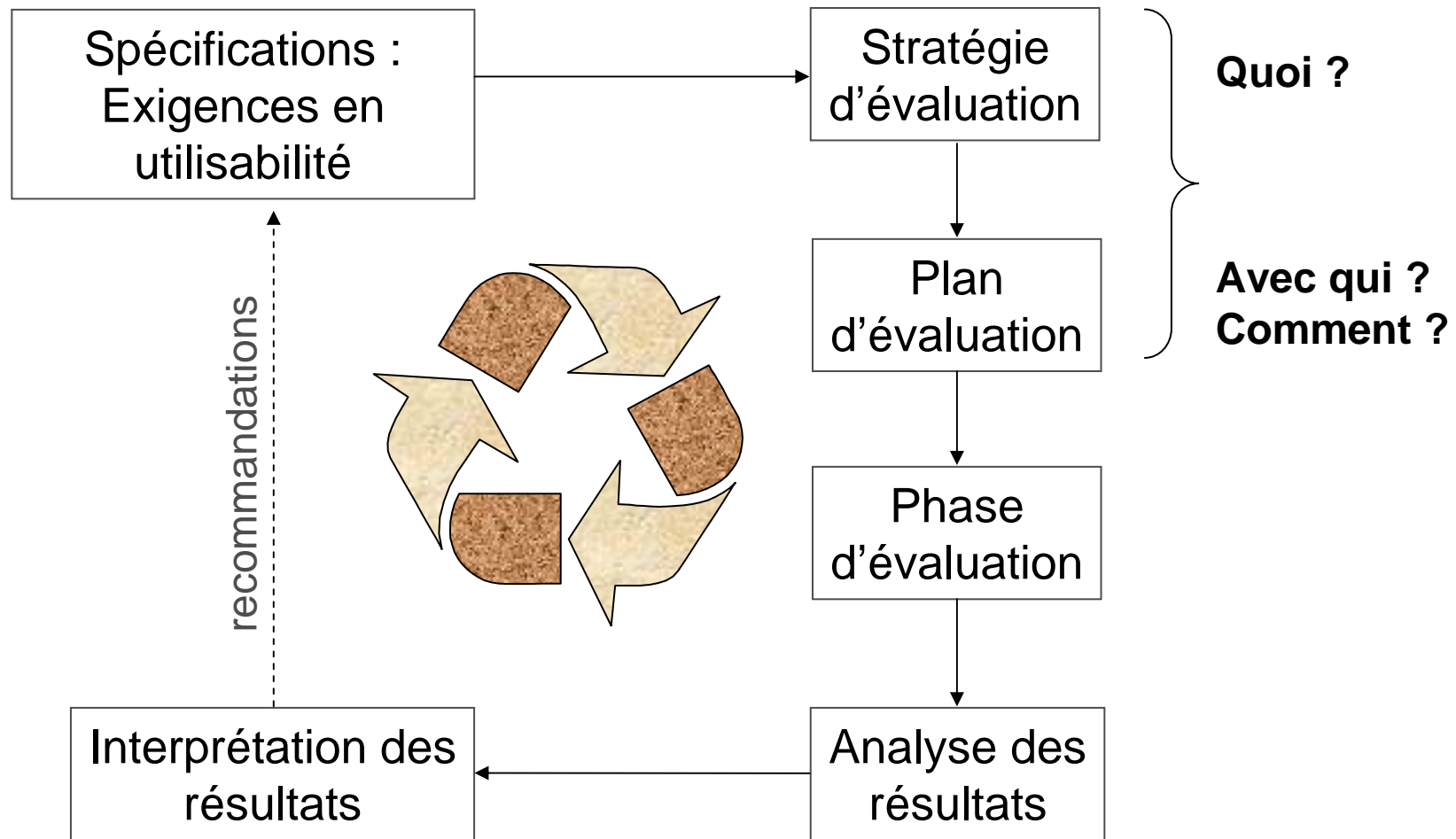
⇒ Stratégie et plan d'évaluation

# EVALUATION DE L'INTERACTION

Comment procéder ?

Processus d'évaluation

[Stone et al. 2005]



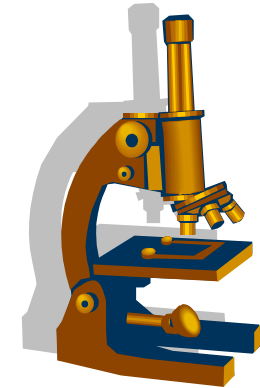
# STRATEGIE D'EVALUATION

## Quoi : que chercher à évaluer ?

- Question principale à se poser : nombreux facteurs d'utilisabilité donc nombreuses façons d'envisager une évaluation.
- Définir des objectifs précis, liés à des facteurs bien identifiés

### ▪ Critères d'utilisabilité de Nielsen

- ✓ *Efficacité* : taux de réussite de l'objectif donné à réaliser
- ✓ *Efficiency* : importance de l'effort cognitif nécessaire
- ✓ *Satisfaction* : ressenti subjectif de l'utilisateur



### ▪ Critères d'utilisabilité de Quesenbery

- ✓ *Efficacité*
- ✓ *Efficiency*
- ✓ *Attrance (caractère engageant)*
- ✓ *Tolérance aux erreurs*
- ✓ *Apprenabilité*

[Quesenbery 2003]



# STRATEGIE D'EVALUATION

## Méthodologie

### 1. Préciser les critères d'utilisabilité importants pour le projet

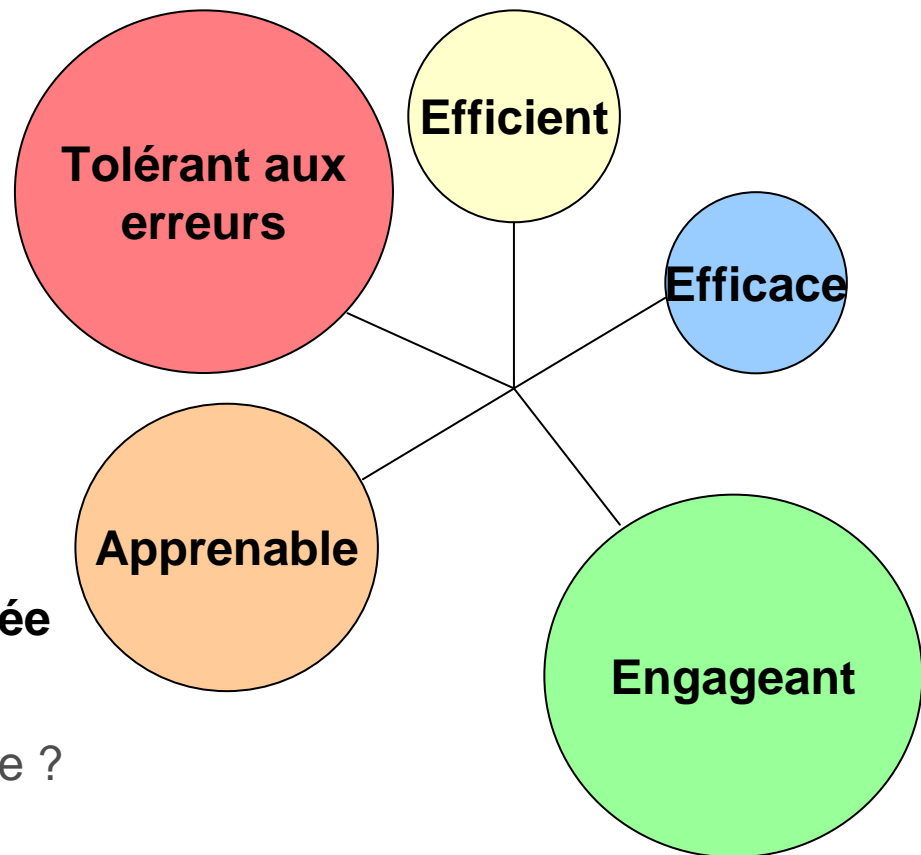
Les spécifications imposées par la MOA peuvent donner des indications sur les attentes en matières d'utilisabilité (public visé, objectifs principaux, voire mesures d'utilisabilité plus précises)

### 2. Hiérarchiser les priorités

Noter l'importance des différents facteurs d'utilisabilité pour le projet (discussion avec le MOA ou notation à partir des spécifications)

### 3. Choisir le type de données adaptées à l'évaluation de chaque critère

Evaluation qualitative ou quantitative ?



# STRATEGIE D'EVALUATION

## Evaluation et recueil de données : deux types d'information

- **Quantitatives** - Basées sur des mesures objectives (durées, % d'erreurs...) obtenues lors des expérimentations.
- **Qualitatives** - Données subjectives non numériques (texte, entretien) obtenues auprès d'utilisateurs ou d'experts en ergonomie des IHMs.

**Spécifications** – Peuvent donner des indications sur le type de données à privilégier. Elles imposent parfois des mesures quantitatives très précises

- *L'animation de la page d'accueil doit s'afficher en moins de 5 sec pour un utilisateur ne disposant que d'une connexion ADSL standard.*
- *Un utilisateur habitué ne doit jamais passer plus de 2 minutes pour saisir les informations clients qui lui sont demandées*
- *Lors de questionnaires post-utilisation, plus de 80% des utilisateurs novices qualifient de « facile d'utilisation » ou » « très facile d'utilisation » la prise en main du logiciel sur une échelle à 5 degrés d'apprenabilité.*

Sinon, c'est à vous de proposer une stratégie d'évaluation que vous pourrez faire valider par le MOA.

# STRATEGIE D'EVALUATION

## Métrique d'évaluation quantitative : norme ISO 9241

Mesures d'utilisabilité définies suivant une grille d'analyse en deux dimensions basée sur des critères ergonomiques étudiés en cours

- ▶ **Efficacité** – atteinte du résultat désiré
  - ▶ **Efficience** – minimisation des ressources (temps, charge cognitive...)
  - ▶ **Satisfaction de l'utilisateur** – ressenti subjectif
- 
- ▶ **Facilité d'apprentissage** – cf. apprenabilité
  - ▶ **Facilité d'appropriation** – apprenabilité liée à la 1<sup>ère</sup> prise en main
  - ▶ **Facilité d'utilisation**
  - ▶ **Fiabilité** – robustesse : erreurs d'utilisation

# STRATEGIE D'EVALUATION

## Métrique d'évaluation quantitative : norme ISO 9241

### Exemple

	Mesures d'utilisabilité		
	Efficacité	Efficience	Satisfaction
Adéquation à la tâche	% de buts atteints	Temps pour réaliser la tâche	Echelle de jugement de satisfaction
Approprié pour l'utilisateur entraîné	Nombre de fonctions importantes utilisées	Efficacité relative comparée à un expert	Echelle de jugement de satisfaction
« Apprenabilité »	% de fonctions apprises	Temps d'apprentissage	Echelle de jugement de la facilité d'utilisation
Tolérance à l'erreur	% d'erreurs corrigées	Temps passé à corriger les erreurs	Echelle de jugement de facilité de correction

**Tableau 2.1.** Mesures de l'utilisabilité préconisées par la norme ISO 9241

# PLAN D'ÉVALUATION

## Stratégie d'évaluation

On sait désormais précisément

- Ce qu'on veut évaluer
- Sur quelles données reposera cette évaluation

## Plan d'évaluation

Il reste de nombreux choix à faire pour mettre en place de manière opérationnelle l'évaluation :

- **Qui** utilisera le système lors de l'évaluation ?
- **Où** (dans quelles conditions) se déroulera l'évaluation ?
- **Quand** (quelle état d'avancement du projet) se déroulera l'évaluation ?

Le plan d'évaluation répond à ces questions en définissant :

- La **technique d'évaluation** la plus appropriée
- La **préparation de l'évaluation** en fonction de cette technique

# PLAN D'ÉVALUATION

## Avec qui évaluer ?

### ✓ Avec des utilisateurs : expérimentation

- Observer les utilisateurs
- Demander leur avis
- Mesurer leurs performances

#### **Recrutement**

- Profil utilisateur (age, expérience...) répondant à la cible
- Nombre de sujets – cinq minimum par groupe de profils
- Rémunération ou non des sujets

### ✓ Sans utilisateur : analyse a priori (ou heuristique)

- Demander leur avis à des experts
- Simuler le fonctionnement en vérifiant des heuristiques ergonomiques
- Estimer le comportement du système à l'aide d'un modèle

**Recrutement – ne jamais faire appel au concepteur du système**

# PLAN D'ÉVALUATION

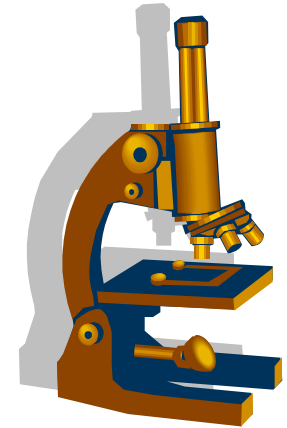
## Paradigmes d'évaluation

### Evaluation analytique / heuristique

- Évaluation a priori
  - revue d'expert (analyse heuristique)
  - promenade cognitive (*cognitive walkthrough*)
  - évaluation participative (*brainstorming*)
- Modèles prédictifs (*loi de Fitts, Keystroke model...*)

### Evaluation expérimentale : objective ou subjective

- Évaluation subjective : questionnaires ou interviews (entretiens)
- Tests d'utilisabilité avec des utilisateurs potentiels
- Tests d'acceptabilité auprès de populations échantillon
- Évaluation post-commercialisation (ou version de test)
- Expérimentations cognitives



# EVALUATION ANALYTIQUE / HEURISTIQUE

## Modèles prédictifs

- ✓ Modèles psycho-cognitifs permettant de simuler les conséquences d'un choix de conception (temps d'exécution...)
- ✓ Exemples : GOMS, Keystroke, Fitts (chapitre IX)

## Evaluation heuristique

[Nielsen & Mack, 1994]

- ✓ Revue du système par plusieurs experts (règles des 5 experts) : simulation d'usages
- ✓ Validation d'un certain nombre d'heuristiques ergonomiques
- ✓ Sur spécifications d'écran et d'interaction (évaluation a priori) ou prototype
- ✓ Plusieurs passes (enchaînement des displays, puis sur chaque display)
- ✓ Doit être validée in fine par des tests utilisateurs sur produit fini



# EVALUATION ANALYTIQUE / HEURISTIQUE

## Evaluation heuristique : mise en œuvre pratique

1. **Définir des tâches** comme pour une évaluation expérimentale avec des sujets naïfs (cf. infra) + **préciser les heuristiques** à étudier
2. **Mettre en place l'évaluation** : pas besoin de conditions d'utilisations réelles
3. **Recueil des données** : l'expert remplit une fiche d'évaluation
4. **Synthèse** : hiérarchiser les problèmes en terme de sévérité et de criticité

## Exemples de fiche d'évaluation heuristique

<b>Tâche n° :</b> <b>Version du système :</b>		<b>Expert :</b> <b>Date :</b>	
Localisation dans la tâche	Heuristique	Description du problème	Commentaires ou suggestion de solution
<i>Un nouveau mail arrive dans la boîte aux lettres</i>	<i>Observabilité</i>	<i>L'utilisateur n'est pas informé de cette arrivée</i>	<i>L'utilisateur aimerait sans doute être alerté (par un son par exemple) de cette arrivée. Option paramétrable.</i>

# EVALUATION ANALYTIQUE / HEURISTIQUE

## Rappel : heuristiques de Nielsen

[Nielsen, 1993]

Bonne base pour une définir une évaluation heuristique

1. **Observabilité** (*visibility of the system status*)
2. **Familiarité** (*match between system and the real world*)
3. **Contrôle** (*user control and freedom*)
4. **Consistance** (*consistency and standards*)
5. **Prévention des erreurs** (*errors prevention*)
6. **Faciliter la remémoration** (*recognition rather than recall*)
7. **Flexibilité et efficacité** (*flexibility and efficiency of use*)
8. **Esthétique** (*aesthetic and minimalist design*)
9. **Gestion des erreurs** (*help users recognize, diagnose and recover from errors*)
10. **Aide** (*help and documentation*)

# EVALUATION ANALYTIQUE / HEURISTIQUE

## Promenade cognitive

[Nielsen & Mack, 1994]

[Spencer,2000]

1. Spécification des utilisateurs visés et du système à réaliser sous forme d'enchaînement d'écrans
2. Évaluation a priori par des experts en présence du concepteur
3. L'évaluateur se « promène » à travers les écrans en simulant la réalisation de la tâche suivant un scénario crédible. Il évalue :
  - ✓ si l'action à réaliser apparaîtrait de manière évidente à l'utilisateur,
  - ✓ si l'utilisateur percevrait aisément que l'action à réaliser est disponible,
  - ✓ si l'utilisateur pourra voir le résultat de son action et l'interprétera correctement
4. Revue critique de l'évaluation avec le concepteur
5. Document de synthèse

# EVALUATION ANALYTIQUE / HEURISTIQUE

## Evaluation participative

[De Bono, 1992]

[Krueger 2000]

### Objectifs

- Réservée aux experts, technique issue du marketing
- Utilisée avant tout en conception initiale pour faciliter la créativité (effet ping-pong entre les partenaires) mais peut aussi concerner l'évaluation

### Protocole

- 1) **Lancement** - Un évaluateur lance le thème qui doit être exploré (ex : « rendre notre site plus visible pour les professionnels »)
- 2) **Production d'idées** – chaque participant lance à tour de rôle des idées qui sont notées les unes après les autres, sans tentative de regroupement et sans qu'un participant ne critique les propositions des autres. Idées visibles de tous (tableau par exemple)
- 3) **Synthèse** : on revient ensuite sur les idées proposées.

## Brainstorming vs. Focus Group

Focus Group plus centré sur la résolution d'un problème précis

# EVALUATION EXPERIMENTALE

## Principe : observation de sessions d'utilisation

Avec ou sans participation visible de l'expérimentateur.

**Intérêt** : conditions d'usages aussi réelles que possible

1. Session d'utilisation du système par un sujet suivant une tâche ou un scénario clairement défini
2. Observation/enregistrement de la session et dépouillement des données.
3. Analyse des données et rédaction d'un rapport d'évaluation

## Différentes analyses

- **Évaluation qualitative** : protocole verbal (*thinking aloud*), questionnaires
- **Évaluation quantitative** : métriques par observation ou fichiers de logs

## Limitations

- Évalue le plus souvent la première prise en main du logiciel : pas de suivi de l'apprentissage au cours du temps
- Ne permet pas une couverture large des fonctionnalités
- Adaptation du comportement : effet Hawthorne

# EVALUATION EXPERIMENTALE

## Où évaluer (tests avec utilisateurs) ?

### ✓ **Sur le terrain**

- Conditions écologiques d'utilisation
- Influence de l'environnement qui peut parasiter l'analyse des données
- Analyse des données moins évidentes, mais permet de détecter des problèmes auquel on n'aurait pas pensé (et qui ne seraient donc pas dans le plan d'expérimentation en laboratoire)

### ✓ **En laboratoire d'utilisabilité**

- Etude plus contrôlée
- Sur une maquette ou des aspects précis
- Récupérer des données objectives plus facilement analysables
- Caractère artificiel qu'il convient de gommer au mieux

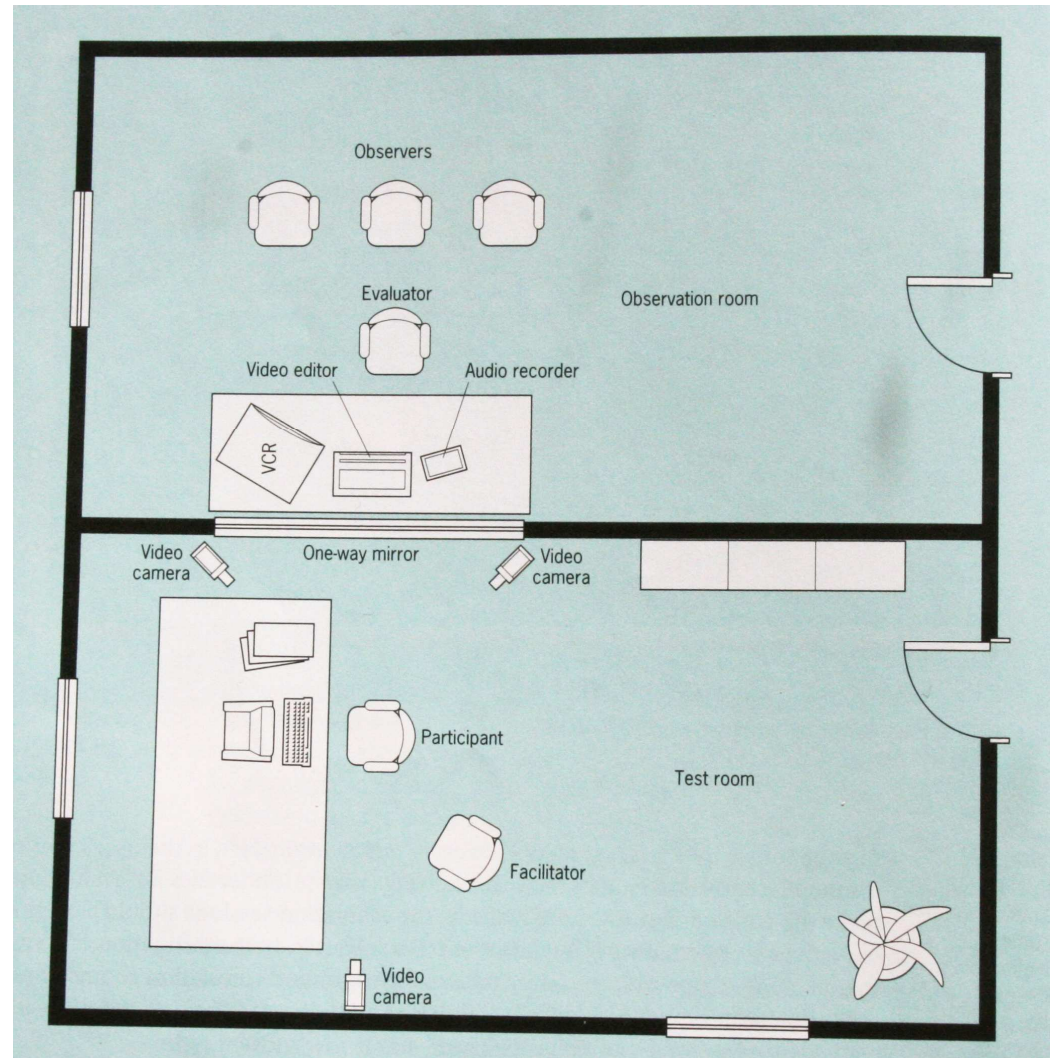
# EVALUATION EXPERIMENTALE

## Laboratoire d'utilisabilité

Salle équipée de tout le matériel permettant d'observer un utilisateur prenant en main le logiciel interactif

### Exemples

- IBM (Boca Raton, Floride),
- Apple
- Microsoft
- Xerox (Palo Alto)
- Sun



# EVALUATION EXPERIMENTALE

## Laboratoire d'utilisabilité

Certaines structures vous permettent de bénéficier des équipements d'un laboratoire d'utilisabilité et d'experts pour conduire vos expériences

**Exemple : Multicom (Grenoble)**

<http://multicom.imag.fr/>



évaluateur /  
observateur

Miroir sans teint

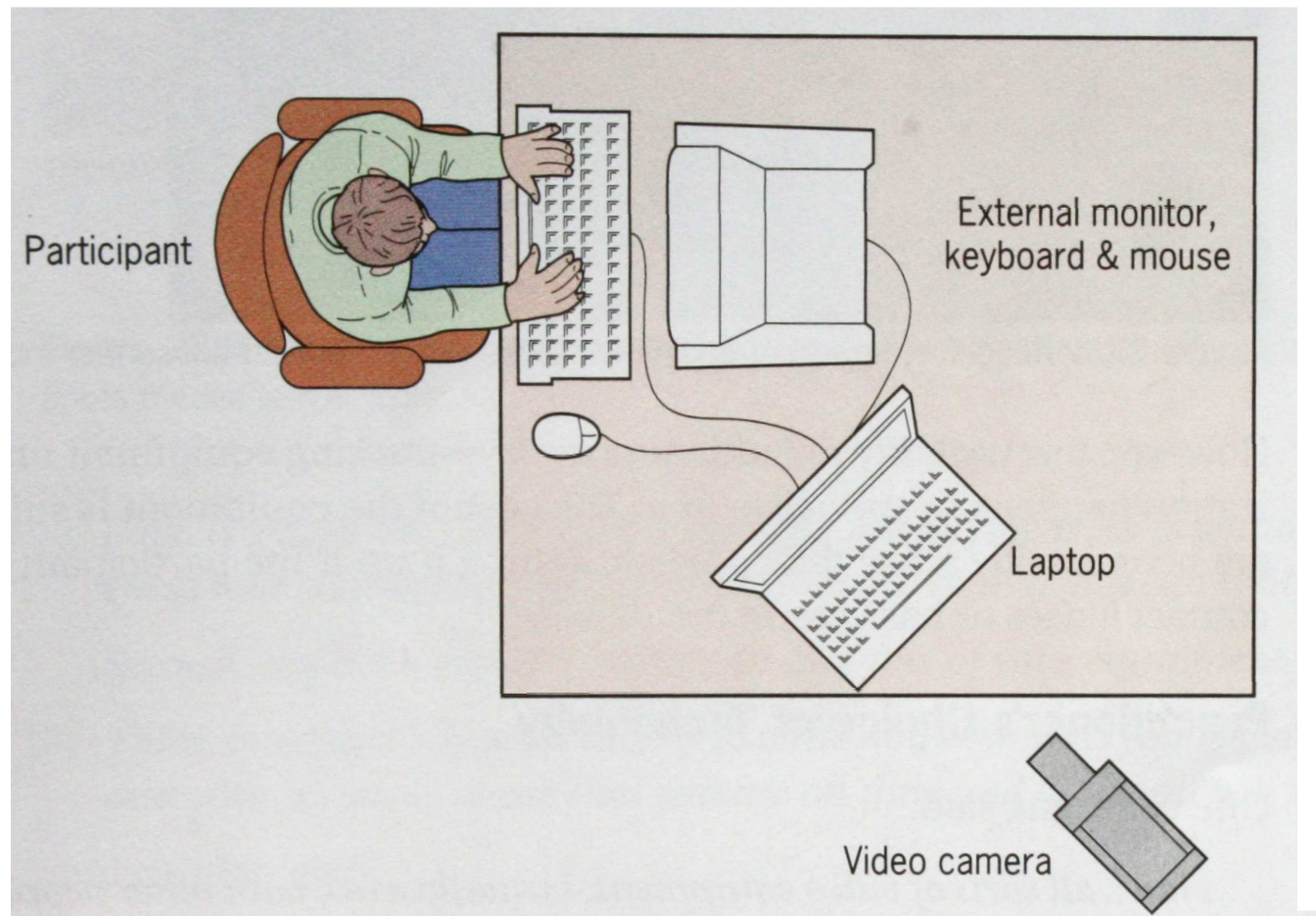




# EVALUATION EXPERIMENTALE

## Laboratoire d'utilisabilité

- Configuration « cheap »

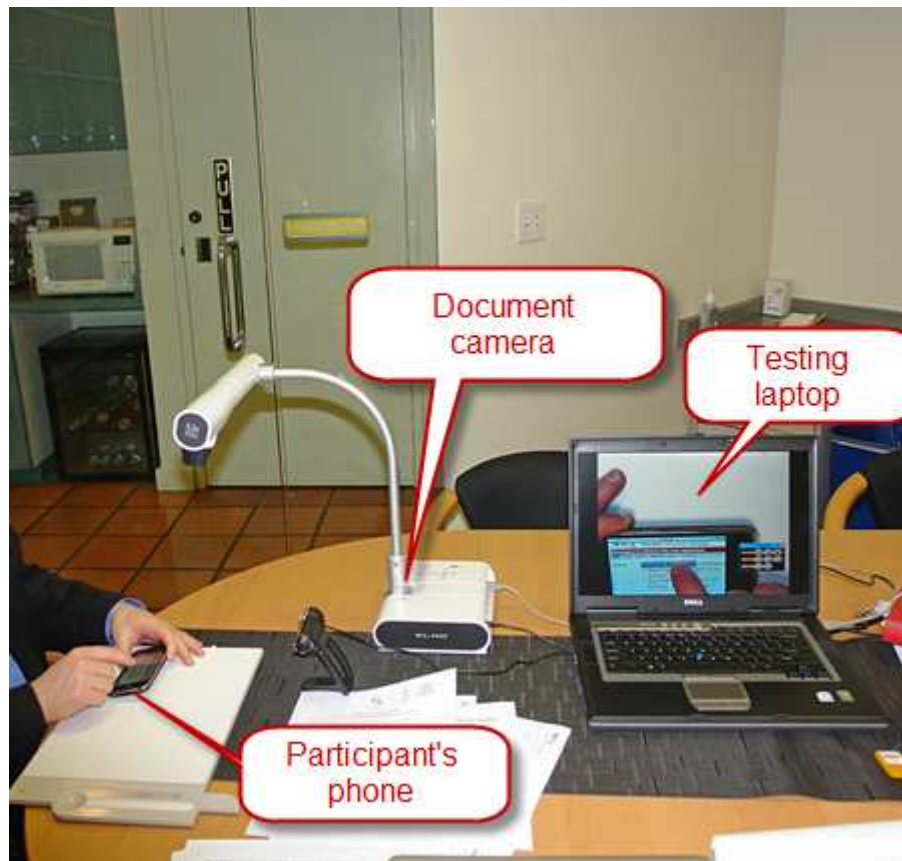


# EVALUATION EXPERIMENTALE

## Laboratoire d'utilisabilité

- Configuration « cheap » interaction sur Mobile

[Budiu 2014]



# EVALUATION EXPERIMENTALE

## Déroulement d'une session d'évaluation : script d'évaluation

Rédiger un **script d'évaluation** qui sera lu au participant permet de nous assurer que le déroulement des sessions est rigoureusement identique (du point de vue de l'utilisateur) d'un sujet à un autre.

### 1. Préparation de la session d'évaluation

- Accueil du sujet
- Recueil éclairé de consentement
- Mise en conditions

### 2. Session d'expérimentation

- Tâche précise (avec ou sans apprentissage préalable)
- Recueil de données et observables en parallèle

### 3. Discussion post-session (facultatif)

- Débriefing : questionnaire post-session
- Analyse rétrospective : visionnage avec le sujet de la session passée



# EVALUATION EXPERIMENTALE

## Déontologie : consentement éclairé

- **Formulaire d'acceptation**

Expliquer au sujet avant la signature du formulaire

- Quel est le but de l'évaluation et ce que l'on attend du sujet
- Quelles sont les informations personnelles qui vont être demandées au sujet : promettre l'anonymat
- Qu'il peut s'arrêter lorsqu'il le souhaite au cours de la session et d'une manière générale quels sont ses droits
- Quels sont les enjeux financiers de l'évaluation (que le sujet soit rémunéré ou non) s'il y en a. De même, préciser s'il y a une clause de confidentialité sur cette expérience.
- Quels sont les risques (limités en IHM !) pris par le sujet lors du test

S'assurer enfin (et seulement à ce moment) de son accord en lui faisant signer un formulaire de consentement.

- **Problème:** évaluation anonyme réalisée sur le WWW

# EVALUATION EXPERIMENTALE

## Préparation de la session d'utilisation

La lecture du script d'évaluation doit permettre au participant de comprendre les objectifs de l'évaluation, la manière dont elle va se dérouler et d'obtenir ainsi son **consentement éclairé**.

### Un exemple de script

*Bonjour, merci d'avoir accepté de participer à ce test. Je vais vous lire un texte qui va vous paraître un peu formel mais qui nous assure que tous les participants recevront la même information, c'est important pour cette étude (...)*

*L'objectif de notre étude est de tester un (...) qui a été conçu pour (...). Nous avons envie de savoir si (...) et nos observations nous permettront d'améliorer ce produit qui sera **commercialisé**. Au cours de cette expérience, je vais vous demander de faire **plusieurs tâches** bien définies. **Ce n'est pas vous que nous testons**, mais le logiciel. Ce test devrait prendre **30 minutes**, mais **vous pourrez l'arrêter** quand vous le voudrez si vous ne vous sentez pas à l'aise (...). Je prendrai des notes tout au cours de l'expérience et vous poserai des questions en fin de test sous forme de questionnaire. J'insiste sur le fait que les données recueillies seront **confidentielles** et que **votre nom ne sera pas enregistré** avec ces informations.*

# EVALUATION EXPERIMENTALE

## Préparation de la session d'utilisation (2)

*Si vous êtes d'accord avec ces conditions, je vous propose de lire ce **formulaire de consentement** et de le signer si vous l'approuvez. Le texte est un peu formel, mais n'hésitez pas à me poser des questions dessus si vous le désirez (...)*

*Merci. Bien, nous pouvons commencer. Dans un premier temps, je vais vous demander de remplir ce questionnaire qui nous permettra de mieux vous connaître. Une fois encore, ces données resteront confidentielles et n'ont pas pour objectif de vous tester, mais de voir à quel profil d'utilisateur vous correspondez (...)*

*Merci, maintenant nous allons commencer la session de test réellement. (...) Vous allez avoir à réaliser plusieurs tâches à la suite. A chaque fois, je vous présenterai une carte où sera écrite la tâche à réaliser. (...) La tâche peut vous sembler difficile à réaliser, cela montrera que le système est mal conçu : vous pouvez donc l'interrompre si vous le souhaitez (...)*  
***Think aloud** : tout au long de l'expérience, je serais heureux si vous commentiez à voix haute vos actions ou ce que vous ressentez (...)*

Le script se poursuit en cas de discussion post-session

# EVALUATION EXPERIMENTALE

## Quoi : quelles tâches tester ?

**Type d'évaluation coûteuse** : impossibilité de tester toutes les fonctionnalités.

**Focalisation** sur quelques tâches, souvent mises en avant par les spécifications

- Tâches les plus fréquentes ou les plus importantes
- Tâches correspondant à des fonctionnalités nouvelles
- Tâches critiques en terme de sécurité
- Tâches mises en avant par le marketing
- ...

## Tâches prédéfinies : intérêt et limites

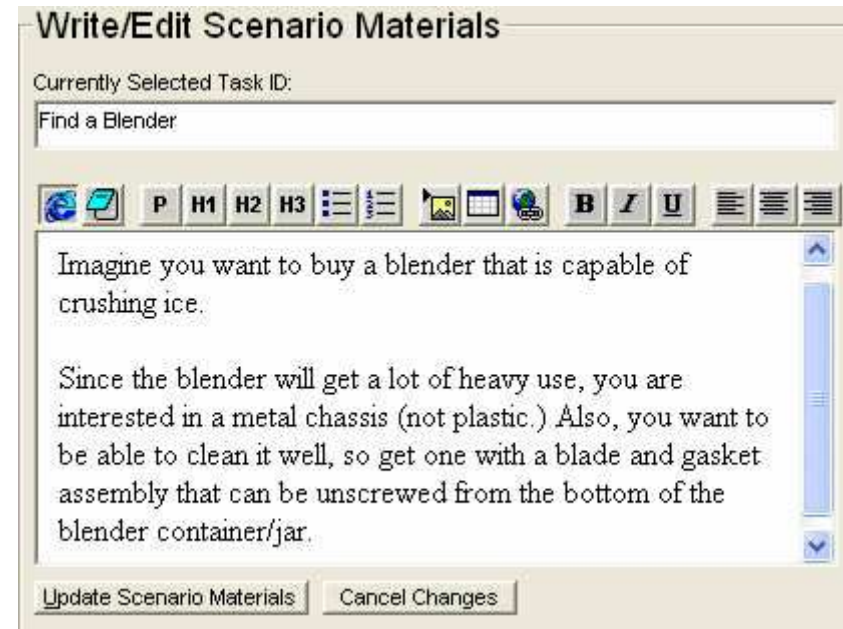
- **Contrôle de l'évaluation** : facilite l'analyse des résultats
- Ne peut vérifier que des situations déjà anticipées, alors qu'on ne peut prévoir le comportement ou les attentes des utilisateurs

**Solutions de compromis** : définir des tâches générales, ou laisser l'utilisateur suggérer une tâche qui lui tient à cœur en fin d'évaluation, quand il connaît bien les fonctionnalités qui lui sont proposées

# EVALUATION EXPERIMENTALE

## Fiche de présentation des tâches

- Fiche papier (préférable)
- Affichage à l'écran : peut interférer avec le logiciel à évaluer



Ovo Logger 6.0

Ovo Logger 7.0

<http://www.ovostudios.com/ovologger.asp>

## Exemples de définition de tâche

*Vous avez vu une publicité pour une promotion en cas d'inscription en ligne à un club de golf et vous êtes intéressé. L'annonce indiquait l'adresse internet du site : [www.golfcheap.com](http://www.golfcheap.com). Retrouvez le formulaire d'inscription adéquat et pré-inscrivez-vous. **Vous pouvez utiliser une fausse identité si vous préférez ne pas donner des informations personnelles vous concernant.***



# EVALUATION EXPERIMENTALE

## Observables

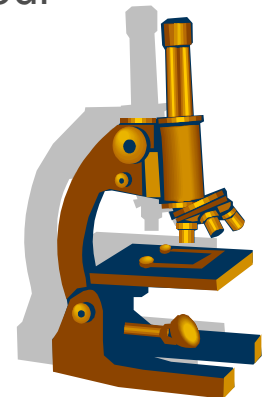
- Postures, direction du regard, productions orales (monologue ou dialogue)
- Actions informatiques, chemin suivi sur un site WWW (fichiers de logs)
- Mesures objectives : durées ou fréquence des actions...

## Un observable particulier : les erreurs

- Essentielles et révélatrices des intentions de l'utilisateur
- Difficiles à observer : l'utilisateur veut généralement les cacher (intérêt d'une capture vidéo / audio)
- Difficiles à analyser : une analyse restrospective avec le sujet permet souvent de mieux comprendre ce qu'il a pensé au moment de l'erreur

## Méthodologie de recueil des observables

- Mesures objectives
- Observation : papier-crayon, analyse de vidéos...



# EVALUATION EXPERIMENTALE

## Recueil manuel d'observation

- Observation : en direct **puis** éventuellement sur vidéo
- Papier-crayon préférable à application informatique
- **Tâche lourde** : prévoir un minimum de soutien matériel, par exemple à l'aide de feuilles pré-remplies.

## Exemples de fiches d'observation pré-définies

Analyse de la succession des actions →

Analyse avec minutage ↓

Temps observables	0-1 min.	1-2 min.	...
Consulte un document			
Consulte l'écran			
Demande conseil à un collègue			
Sélectionne « Insertion Graphique »			

	Recherche nom	Sélectionne nom	Retour	Volume	Appel téléphonique	Haut-parleur
Rechercheur	x					
		x				
			x			
	x					
		x				
					x	

# EVALUATION EXPERIMENTALE

## Exemples de fiche de recueil d'observations :

Fiche plus générale dans le cas d'une évaluation en *think aloud*

<b>Tâche n° :</b>		<b>Sujet n° :</b>
<b>Nom de l'évaluateur :</b>		
<b>Version du système :</b>		<b>Date :</b>
Action mise en jeu dans la description de la tâche	Remarque du participants	Commentaires / explications

## Recueil automatique de données d'observation

- Utilitaires de recueil mais également d'analyse de logs

**Exemple :** Ovo Logger

[www.ovostudios.com/ovologger.asp](http://www.ovostudios.com/ovologger.asp)

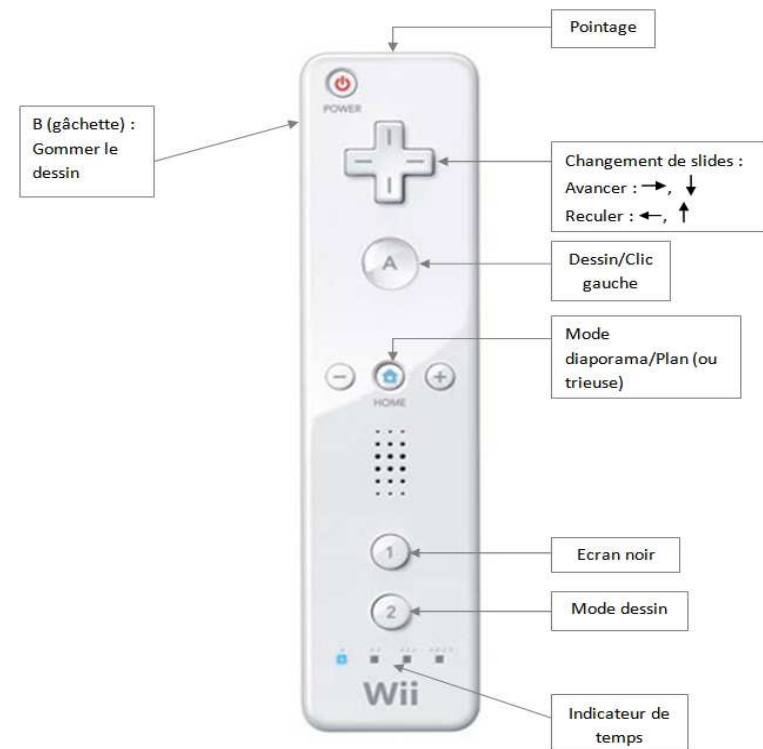
# EVALUATION EXPERIMENTALE

## Exemple : commande vidéoprojecteur avec WiiMote

Perrin / Raynard  
U. Nice (2007-2008)



❶ Dispositif expérimental



❷ Tutoriel

# EVALUATION EXPERIMENTALE

## Exemple : commande vidéoprojecteur avec WiiMote

Gestes	Avancer d'un slide	Reculer d'un slide	Mode plein écran / Mode plan	Lancer un hyperlien (pointage)	Ecran noir	Dessiner (pointage)	Surligner (pointage)	Zoom +	Zoom -
Sujet 1				P + (A)		(1)	(B)	(A) +	(A) +
Sujet 2 G	(B) +	(B) +	↑ / ↓	(A) ou (B)	(1)	(A) ou (B)	(A) ou (B)	(A) ou (B)	(A) ou (B)
Sujet 3			(A)			(B)	(1)	(+) +	(+) +
Sujet 4	→	←	1/2 +	(-)	(B)	(A)	(A)	↑ +	↓ +
Sujet 5				(A)	(1)	(B)	(2)	(+)	(-)
Sujet 6				(A) ou (B)	(1)	(A) ou (B)	(A) ou (B)	(A) ou (B) +	(A) ou (B) +
Sujet 7			(1)	(A)	(2)	(A)	(A)	↑	↓

### ③ Observations et analyse

# EVALUATION EXPERIMENTALE

## Autres observables : parole

[Hix & Hartson 1993]

**Protocole *Think aloud*** – Le sujet exprime ce qu’il ressent durant le test : productions libres ou réponse à des sollicitations/questions de l’expérimentateur

- *Intérêt* – Elicitation claire des problèmes rencontrés
- *Limitation* – peut le perturber la réalisation de la tâche

## Autres observables : suivi du regard (*eyetracking*)

Solutions à base de caméra infra-rouge



**Tobii 1750**  
[www.tobii.com](http://www.tobii.com)



**Mobile Eye (ASL)**  
[www.a-s-l.com](http://www.a-s-l.com)



# EVALUATION EXPERIMENTALE

## Dépouillement

- **Evaluation qualitative** – Problèmes les plus flagrants : cas exemplaires
- **Evaluation quantitative** – Calcul de mesures synthétiques à partir des données d'observation + analyse statistique de données.

## Exemples de mesures

[Whiteside, Bennett et Holtzblatt 1988]

- ✓ Temps d'exécution d'une tâche donnée
- ✓ % de la tâche exécutée complètement
- ✓ Ratio sessions réussies / échecs
- ✓ Distribution du nombre d'erreurs suivant les sujets
- ✓ Temps perdu sur des erreurs
- ✓ Nombre de commandes utilisées pour réaliser la tâche
- ✓ Fréquence d'utilisation de l'aide ou de la documentation
- ✓ % de commentaires favorables / défavorables (*think aloud*)
- ✓ Nombre de répétitions d'une commande erronée
- ✓ Nombre de commandes invoquées non utilisées
- ✓ Nombre de fois où le sujet a perdu le contrôle du système
- ✓ Nombre de fois où le sujet exprime une frustration

# EVALUATION EXPERIMENTALE

## Dépouillement : échantillonnage de la population de test

- **Diversité des utilisateurs** : importance de l'échantillonnage de la population quelque soit la technique d'expérimentation choisie (sauf *quick and dirty*...)
- **Plan d'évaluation**
  1. Caractériser les profils d'utilisateurs visés (sexe, age, expert / novice, maîtrise de l'outil informatique, catégorie socio-professionnelle...)
  2. Échantillonner la population suivant les profils retenus socioprofessionnelle...

Taille de l'échantillon : 5, 12, 20, 100?

[Dumas & Reddish, 1999]

- **Dépouillement et analyse des résultats**

- Analyse multi-critères suivants les différentes caractéristiques de profils
- Pertinence statistique des résultats
- **Discipline à part entière** : statistique (*protocoles et tests*)



# EVALUATION ET DIVERSITE D'UTILISATEURS

## Dépouillement : analyse statistique des résultats

	1	2	3	4	5	6	7	Moy
Age	37	41	43	54	46	44	61	<b>40.9</b>
Sexe	F	F	M	M	F	F	M	<b>40,3</b>
Niveau d'étude	4	2	4	4	4	1	2	<b>3.0</b>
Année PC	5	2	0	2	6	4	9	<b>4.0</b>
Facilité usage	1	2	2	1	2	3	1	<b>1.7</b>
Qualité de l'aide	1	3	3	1	3	2	2	<b>2.1</b>



**Facilité d'usage**

25-50 ans



2 / 3

> 50 ans



1.5 / 3

# EVALUATION EXPERIMENTALE

## Synthèse : rapport d'évaluation

- **Evaluation qualitative** – Cas problématiques hiérarchisés avec copies d'écrans, vidéos ou explications...
- **Evaluation quantitative** – présentation des métriques souvent présentes dans les spécifications ou le cahier de recettes

### Exemple

norme ISO 9241-11

<b>Fiabilité</b>	✓ % de buts réalisés
<b>adéquation à la tâche</b>	✓ Temps de réalisation de la tâche
<b>Adaptation aux utilisateurs entraînés</b>	✓ Nb de fonctionnalités avancées utilisées ✓ Efficacité relative avec un expert
<b>Apprenabilité</b>	✓ % de fonctions apprises après un tps t ✓ Temps pour apprendre une fonction
<b>Robustesse : tolérance aux erreurs</b>	✓ % erreurs corrigées ✓ Temps perdu dues aux erreurs

- Dans l'idéal : rapport combinant aspects qualitatifs et quantitatifs

# EVALUATION EXPERIMENTALE

## Rapport d'évaluation : tests d'acceptation

- Très fréquemment employé pour le système final : spécification
- Validation dépendant de **seuils de réussite** attendus sur les résultats

Après 5 heures d'utilisation par des novices et 15 jours d'attente, 50% de la population de test doit être capable de réaliser 75% de la tâche de test correctement.

- **Attention à la définition des seuils** – Exemple : pré-tests sur ancien logiciel

**Exemple** : norme ISO 9241-11

Basé sur les facteurs d'utilisabilité de Nielsen

- ✓ **Efficacité** : vérification que les objectifs visés par l'utilisateur sont atteints
- ✓ **Efficience** : évaluation des ressources mises en œuvre pour atteindre cet objectif (par exemple, temps pour réaliser une tâche)
- ✓ **Satisfaction** : quantification du niveau de satisfaction de l'utilisateur



- ✓ **Efficacité** : OK si 95 % des sujet réussissent à accomplir la tâche
- ✓ **Efficience** : OK si 90% des sujet réalisent une tâche en un temps donné
- ✓ **Satisfaction** : OK si moins de 10% des sujets expriment un problème de fonctionnement

# EVALUATION SUBJECTIVE

## Principe : opinion post-utilisation

1. Session d'utilisation, généralement suivant une tâche ou un scénario clairement définis
2. Recueil post-session de l'avis des sujets : questionnaire, interview ou visionnage de la vidéo de la session d'utilisation
3. Synthèse qualitative ou quantitative suivant les données recueillies : peut donc faire partie intégrante d'une évaluation expérimentale objective.

## Intérêt



- Point de vue de l'utilisateur + explications sur les difficultés rencontrées
- Evaluation moins contrôlée et plus aisée à mettre en œuvre : plus couvrante
- **Problème** : données subjectives parfois différentes à analyser

## Différentes techniques

- **Interview** libre ou orienté ⇨ évaluation qualitative
- **Questionnaires**: échelles de valeurs sur des points précis ⇨ **quantitatif**

# EVALUATION SUBJECTIVE

## Interview libre

- Le sujet aborde des points qui l'ont marqué et qui n'avaient peut-être pas retenu l'attention du concepteur 
- Manque d'homogénéité des avis, précision variable: analyse synthétique difficile 

## Interview dirigé / semi-structuré

[Nielsen et al. 1986]

- Questions ouvertes ou fermées testant des points précis : la liste des questions doit être définie avant la session d'évaluation
- Évaluation structurée: analyse facilitée

- |   |   |
|---|---|
| ✓ Aviez-vous déjà réservé un hôtel en ligne?            | <input type="checkbox"/> oui <input type="checkbox"/> non |
| ✓ Cette fonctionnalité vous semble-t-elle intéressante? | <input type="checkbox"/> oui <input type="checkbox"/> non |
| ✓ Êtes-vous facilement arrivé à réserver un hôtel?      | <input type="checkbox"/> oui <input type="checkbox"/> non |
| ✓ Cela vous a-t-il pris trop de temps?                  | <input type="checkbox"/> oui <input type="checkbox"/> non |

# EVALUATION SUBJECTIVE

## Interview semi-structuré : exemple

[Nielsen et al. 1986]

Idées de questions pour une évaluation rétrospective : visionnage avec le sujet de la vidéo de la session d'utilisation

- ✓ *Why do you do this ?* (connaître le but de l'utilisateur)
- ✓ *How do you do it ?*  
(récupérer les sous tâches pour appliquer récursivement les questions)
- ✓ *Why do you not do this in the following manner ?*  
(comprendre les choix de l'utilisateur)
- ✓ *What are the préconditions for doing this ?*  
(évaluer si l'utilisateur a compris les conditions de déclenchement)
- ✓ *What are the results of doing this ?*
- ✓ *Do errors ever occur when doing this ?*
- ✓ *How do you discover and correct these errors ?*

# EVALUATION SUBJECTIVE

## Questionnaires

- L'utilisateur a parfois du mal à donner des avis tranchés
- Évaluation subjective : échelle multivaluée, ou de Lickert de préférence
- Synthèse finale : analyse statistique de données

## Exemple

Évaluez de 1 (pas du tout) à 4 (tout à fait) votre accord avec ces affirmations :

- |   |                            |                            |                            |                            |
|---|----------------------------|----------------------------|----------------------------|----------------------------|
| ✓ Cette fonctionnalité est intéressante     | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 |
| ✓ Il est facile de réserver avec le système | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 |
| ✓ Le temps de réservation est acceptable    | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 |

## Nombreuses bases de test ... payantes pour la plupart

- **SUS** (System Usability Scale) [Brooke, 1996]
- **QUIS** (Questionnaire for User Interaction Satisfaction) [Chin *et al* 1988]
- **CSUQ** (IBM Computer System Usability Questionnaire) [Lewis 1995]
- **SUMI** : Software Usability Measurement Inventory [Kirakowski et Corbett 1993]



## QUIS (Questionnaire for User Interaction Satisfaction)

Site WWW – [www.lap.umd.edu/QUIS/](http://www.lap.umd.edu/QUIS/)

- Un des questionnaires les plus utilisés
  - Libre sous convention d'utilisation
  - Couvre de nombreux types d'expérimentations
- ✓ expérience passée sur le système testé
  - ✓ expérience passée sur d'autres systèmes
  - ✓ opinion générale de l'utilisateur sur le système
  - ✓ affichage
  - ✓ terminologie utilisée et information fournie par le système
  - ✓ apprenabilité
  - ✓ documentation papier et aide en ligne
  - ✓ documentation en ligne
  - ✓ multimédia
  - ✓ téléconférence et travail collaboratif
  - ✓ installation du système



# EVALUATION SUBJECTIVE



## QUIS : exemples de questions

OVERALL REACTION TO THE SOFTWARE		0	1	2	3	4	5	6	7	8	9		NA
1. <input type="checkbox"/>	terrible	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	wonderful	<input type="radio"/>
2. <input type="checkbox"/>	difficult	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	easy	<input type="radio"/>
3. <input type="checkbox"/>	frustrating	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	satisfying	<input type="radio"/>
4. <input type="checkbox"/>	inadequate power	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	adequate power	<input type="radio"/>
5. <input type="checkbox"/>	dull	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	stimulating	<input type="radio"/>
6. <input type="checkbox"/>	rigid	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	flexible	<input type="radio"/>
SCREEN		0	1	2	3	4	5	6	7	8	9		NA
7. Reading characters on the screen <input type="checkbox"/>	hard	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	easy	<input type="radio"/>
8. Highlighting simplifies task <input type="checkbox"/>	not at all	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	very much	<input type="radio"/>
9. Organization of information <input type="checkbox"/>	confusing	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	very clear	<input type="radio"/>
10. Sequence of screens <input type="checkbox"/>	confusing	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	very clear	<input type="radio"/>
TERMINOLOGY AND SYSTEM INFORMATION		0	1	2	3	4	5	6	7	8	9		NA
11. Use of terms throughout système <input type="checkbox"/>	inconsistent	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	consistent	<input type="radio"/>
12. Terminology related to task <input type="checkbox"/>	never	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	always	<input type="radio"/>
13. Position of messages on screen <input type="checkbox"/>	inconsistent	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	consistent	<input type="radio"/>
14. Prompts for input <input type="checkbox"/>	confusing	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	clear	<input type="radio"/>
15. Computer informs about its progress <input type="checkbox"/>	never	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	always	<input type="radio"/>
16. Error messages <input type="checkbox"/>	unhelpful	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	helpful	<input type="radio"/>
LEARNING		0	1	2	3	4	5	6	7	8	9		NA
17. Learning to operate le système <input type="checkbox"/>	difficult	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	easy	<input type="radio"/>
18. Exploring new features by trial and error <input type="checkbox"/>	difficult	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	easy	<input type="radio"/>
19. Remembering names and use of commands <input type="checkbox"/>	difficult	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	easy	<input checked="" type="radio"/>
20. Performing tasks is straightforward <input type="checkbox"/>	never	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	always	<input type="radio"/>
21. Help messages on the screen <input type="checkbox"/>	unhelpful	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	helpful	<input type="radio"/>

commentaires  
éventuels

# EVALUATION SUBJECTIVE



## SUMI (Software Usability Measurement Inventory)

Site WWW – <http://sumi.ucc.ie/> (gratuit pour usage académique)

**Exemple** : caractérisation du profil utilisateur

Company

Job Title

Age range (*please check one*)

Sex (*please check one*)

Name of software being evaluated

- How long have you worked with this software (*please check one*)
- In a typical week, do you use the software for (*please check one*)
- Describe briefly the sorts of task you carry out with this software
- Were you involved in choosing this software (*please check one*)
- When have you received any training with this software (*please check one*)
- Did you find the training adequate (*please check one*)
- Any further comments on the training?
- For how long have you been using computers in your work

...



## SUMI : questionnaire post-utilisation

- 50 questions classées par facteur d'utilisabilité
- L'utilisateur doit répondre avec une échelle trivaluée : d'accord, sans avis ou incertain, pas d'accord

### Exemples de questions

- **Apprenabilité**

*Je n'arriverai jamais à retenir toutes les fonctions du logiciel*

- **Aide**

*Les instructions et les affichages aident à la compréhension*

- **Contrôle**

*Souvent, je n'ai pas su ce qu'il fallait faire ensuite avec le logiciel*

- **Gestion des erreurs**

*Lors d'un blocage, il n'est pas facile de revenir en arrière à ce que je suis en train de faire*

- **Affect**

*Je recommanderais ce logiciel à un collègue*

# EVALUATION SUBJECTIVE



## QCUS (IBM - Computer System Usability Questionnaire)

WWW – <http://oldwww.acm.org/perlman/question.cgi?form=CSUQ> (n'est plus maintenu)

		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"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Accord	<input type="radio"/>
2. Ce système est simple à utiliser ☑	Désaccord	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Accord	<input type="radio"/>
3. J'ai complété mon travail correctement en utilisant ce système ☑	Désaccord	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<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"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Accord	<input type="radio"/>
5. J'ai complété mon travail efficacement en utilisant ce système ☑	Désaccord	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Accord	<input type="radio"/>
6. Je me sens à l'aise avec ce système ☑	Désaccord	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Accord	<input type="radio"/>
7. J'ai eu de la facilité à apprendre comment utiliser ce système ☑	Désaccord	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Accord	<input type="radio"/>
8. Je crois être devenu(e) rapidement efficace en utilisant ce système ☑	Désaccord	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<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"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<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"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<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"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Accord	<input type="radio"/>
12. J'ai facilement trouvé l'information que je cherchais ☑	Désaccord	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Accord	<input type="radio"/>
13. L'information fournie avec ce système est facile à comprendre ☑	Désaccord	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<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"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Accord	<input type="radio"/>
15. L'organisation de l'information dans les écrans du système est claire ☑	Désaccord	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Accord	<input type="radio"/>
16. L'interface de ce système est agréable ☑	Désaccord	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Accord	<input type="radio"/>
17. J'aime utiliser l'interface de ce système ☑	Désaccord	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<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"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Accord	<input type="radio"/>
19. En général, je suis satisfait(e) de ce système ☑	Désaccord	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Accord	<input type="radio"/>
		1	2	3	4	5	6	7		NA

# EVALUATION SUBJECTIVE



## Microsoft Product Reactions Cards

Approche un peu particulière  
(Qualification par mots)

<input type="checkbox"/> Convenient	<input type="checkbox"/> Familiar	<input type="checkbox"/> Slow	<input type="checkbox"/> Cutting Edge	<input type="checkbox"/> Friendly
<input type="checkbox"/> Busy	<input type="checkbox"/> Straight forward	<input type="checkbox"/> Personal	<input type="checkbox"/> Confusing	<input type="checkbox"/> Stressful
<input type="checkbox"/> Fun	<input type="checkbox"/> Boring	<input type="checkbox"/> Innovative	<input type="checkbox"/> Helpful	<input type="checkbox"/> Simplistic
<input type="checkbox"/> Empowering	<input type="checkbox"/> Usable	<input type="checkbox"/> Old	<input type="checkbox"/> Complex	<input type="checkbox"/> Irrelevant
<input type="checkbox"/> Dated	<input type="checkbox"/> Dull	<input type="checkbox"/> Advanced	<input type="checkbox"/> Patronizing	<input type="checkbox"/> Meaningful
<input type="checkbox"/> Clear	<input type="checkbox"/> Flexible	<input type="checkbox"/> Sophisticated	<input type="checkbox"/> Effective	<input type="checkbox"/> Difficult
<input type="checkbox"/> Accessible	<input type="checkbox"/> Time saving	<input type="checkbox"/> Business-like	<input type="checkbox"/> Too Technical	<input type="checkbox"/> Intuitive
<input type="checkbox"/> Organized	<input type="checkbox"/> Calm	<input type="checkbox"/> Disruptive	<input type="checkbox"/> Fast	<input type="checkbox"/> Consistent
<input type="checkbox"/> Controllable	<input type="checkbox"/> Annoying	<input type="checkbox"/> Easy to use	<input type="checkbox"/> Met Expectations	<input type="checkbox"/> Powerful
<input type="checkbox"/> Novel	<input type="checkbox"/> Comfortable	<input type="checkbox"/> Approachable	<input type="checkbox"/> Collaborative	<input type="checkbox"/> Distracting
<input type="checkbox"/> Frustrating	<input type="checkbox"/> Attractive	<input type="checkbox"/> Efficient	<input type="checkbox"/> Ordinary	<input type="checkbox"/> High Quality
<input type="checkbox"/> Valuable	<input type="checkbox"/> Gets in the way	<input type="checkbox"/> Fragile	<input type="checkbox"/> Connected	<input type="checkbox"/> Overbearing
<input type="checkbox"/> Sterile	<input type="checkbox"/> Exciting	<input type="checkbox"/> Desirable	<input type="checkbox"/> Overwhelming	<input type="checkbox"/> Secure
<input type="checkbox"/> Predictable	<input type="checkbox"/> Intimidating	<input type="checkbox"/> Appealing	<input type="checkbox"/> Clean	<input type="checkbox"/> Understandable
<input type="checkbox"/> Low Maintenance	<input type="checkbox"/> Comprehensive	<input type="checkbox"/> Unrefined	<input type="checkbox"/> Unconventional	<input type="checkbox"/> Effortless
<input type="checkbox"/> Inviting	<input type="checkbox"/> Entertaining	<input type="checkbox"/> Confident	<input type="checkbox"/> Optimistic	<input type="checkbox"/> Enthusiastic
<input type="checkbox"/> Customizable	<input type="checkbox"/> Time consuming	<input type="checkbox"/> Stable	<input type="checkbox"/> Unattractive	<input type="checkbox"/> Stimulating
<input type="checkbox"/> Not valuable	<input type="checkbox"/> Relevant	<input type="checkbox"/> Inconsistent	<input type="checkbox"/> Compelling	<input type="checkbox"/> Reliable
<input type="checkbox"/> Professional	<input type="checkbox"/> Compatible	<input type="checkbox"/> Incomprehensible	<input type="checkbox"/> Useful	<input type="checkbox"/> Undesirable
<input type="checkbox"/> Not secure	<input type="checkbox"/> Trustworthy	<input type="checkbox"/> Integrated	<input type="checkbox"/> Disconnected	<input type="checkbox"/> Energetic
<input type="checkbox"/> Engaging	<input type="checkbox"/> Motivating	<input type="checkbox"/> Poor quality	<input type="checkbox"/> Impressive	<input type="checkbox"/> Unpredictable
<input type="checkbox"/> Hard to use	<input type="checkbox"/> Ineffective	<input type="checkbox"/> Creative	<input type="checkbox"/> Uncontrollable	<input type="checkbox"/> Fresh
<input type="checkbox"/> Responsive	<input type="checkbox"/> Essential	<input type="checkbox"/> Satisfying	<input type="checkbox"/> Exceptional	
<input type="checkbox"/> Impersonal	<input type="checkbox"/> Inspiring	<input type="checkbox"/> Unapproachable	<input type="checkbox"/> Rigid	

# EVALUATION SUBJECTIVE

Un exemple : sites financiers

2 tâches, 123 sujets

[Tullis, Stetson 2004]

The screenshot displays the Kiplinger.com website interface. At the top, there is a navigation bar with the Kiplinger.com logo, a search bar, and a date stamp of June 5, 2004, 2:57 pm. The main content area is divided into several sections:

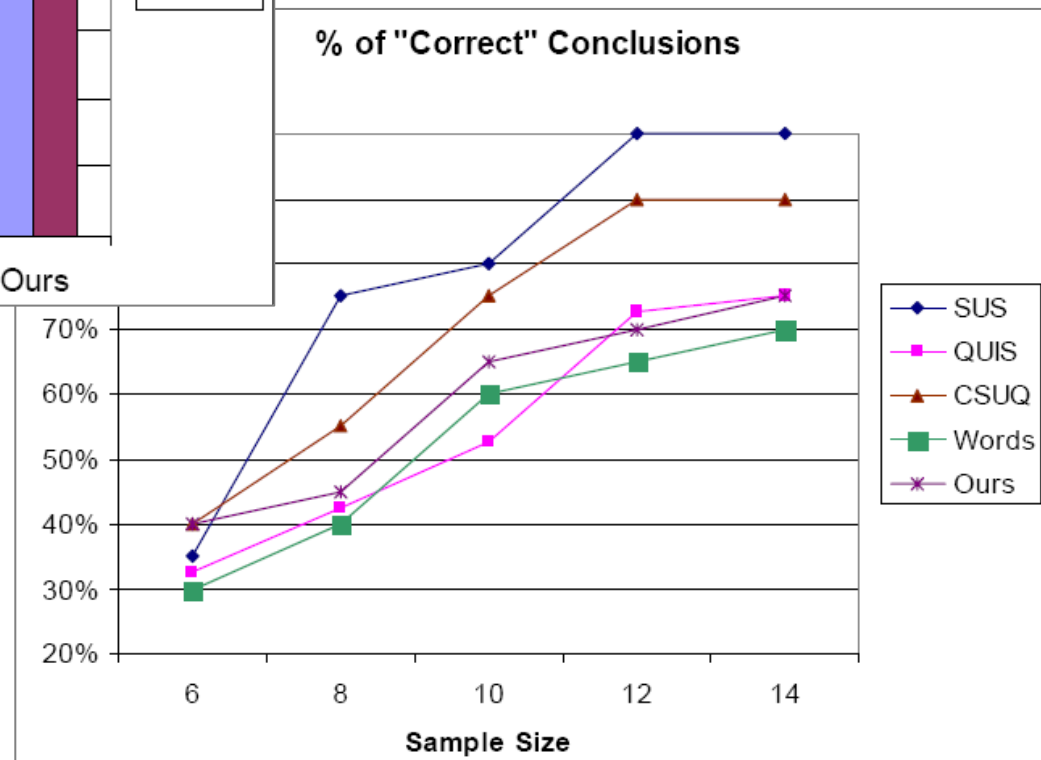
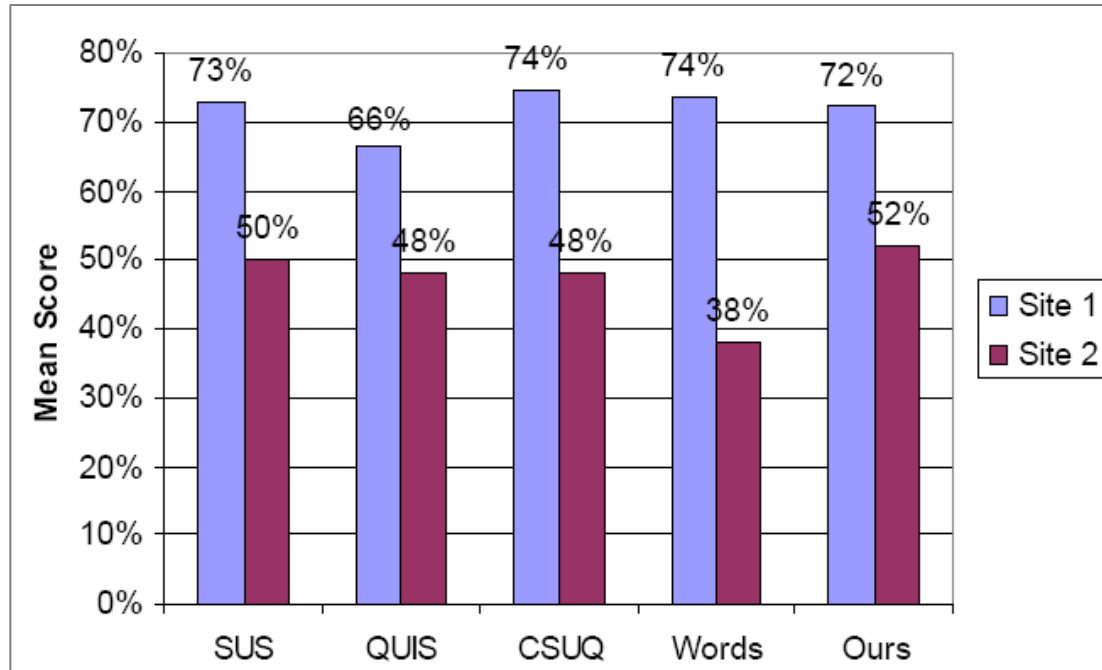
- PERSONAL FINANCE ADVICE:** Features an article titled "Bargain Voyage" by Sean O'Neil, discussing family summer vacation planning. It includes a "Recent Online Features" list with links like "Trend Spotting" and "Gas Pain Relief".
- BUSINESS FORECASTS:** Promotes "The Kiplinger Letter" with a "This Week In" section listing topics like "States See Revival As Revenues Rise" and "New Tax Twist: Private Collectors".
- Market Summary:** A sidebar on the left provides a line graph for Nasdaq 4-Jun 4 4:00pm (CST) and lists market data for Dow, Nasdaq, S&P 500, and 10-Yr Bond.
- Navigation and Services:** Includes a "Money Manager" section, "Top Stocks" (Goldilocks), and "Market Outlook" (Market O).
- Additional Content:** A "Magazine" section, "LATEST UPDATE" (Strong Jobs Data Point to Rate Hike), and "OUTLOOK" (Economic Growth GDP up 4.3% in '04) are also visible.

The interface is cluttered with numerous links, advertisements, and data points, reflecting the complexity of financial websites mentioned in the text.

# EVALUATION SUBJECTIVE

## Résultats

[Tullis, Stetson 2004]



# EVALUATION SUBJECTIVE

## Construire son propre questionnaire : méthodologie

- **Importance de l'ordre des questions**
  1. Questions générales et simples (mise en confiance)
  2. Questions précises et complexes
  3. Questions dépendant du profil utilisateur (en dernier car parfois jugées indiscrètes)
- **Importance de la rédaction des questions**, de l'ordre des réponses
- **Importance de l'échelle de notation** – plusieurs échelles envisageables
  - ✓ **Oui / Non**
  - ✓ **Echelle de Lickert**: *totalemment d'accord, d'accord, moyennement d'accord, sans avis, en désaccord...*
  - ✓ **Ordre d'importance** : *classez ces différentes fonctions suivant leur importance*
  - ✓ ...



# EVALUATION THEORIQUE : MODELES PREDICTIFS

## Modèles prédictifs

- Mesure prédictive du comportement de l'utilisateur sans test réel
- Basés sur des résultats *réels* en psychologie cognitive
- Utile en conception et évaluation

## GOMS

[Card, Moran, Newell, 1983]

- **Goals** — But que l'utilisateur veut réaliser
  - **Methods** — Procédures (séquences d'actions ou activités cognitives) requises pour réaliser le but
  - **Operators** — Processus cognitifs et actions physiques mis en jeu au final par les différentes méthodes envisageables
  - **Selection** — Règles utilisées pour choisir une méthode à un moment donné
- ⑤ Estimation des temps d'exécution à partir de temps moyens connus pour chaque opérateur

# MODELES PREDICTIFS: GOMS (exemple)

- **But** Editeur de texte: supprimer un mot dans une phrase
- **Méthodes** Méthode par sélection de menu
  1. Se rappeler que le mot doit être sélectionné
  2. Se rappeler que la commande est «Couper»
  3. Se rappeler que la commande est dans le menu «Edition»
  4. Accomplir la sélection du mot et de la commande «Couper»
  5. Retour sur fin de tâcheMéthode avec la touche « Supprimer »
  1. Se rappeler que le curseur doit être en fin de mot
  2. Se rappeler quelle est la touche «Supprimer»
  3. Presser la touche «Supprimer» autant que nécessaire
  4. Retour sur fin de tâche
- **Opérateurs**
  - Clic souris
  - Sélection à la souris d'une partie de texte
  - Déplacement curseur à la position recherchée
  - Pression touche clavier
- **Sélection** Règle: menu si texte à ôter long, touche supprimer sinon

# MODELES PREDICTIFS: KEYSTROKE

## Keystroke level model

[Card, Moran, Newell, 1983]

- Indépendant de GOMS
- Estimation du temps d'exécution de commandes de base
- Modèle basé sur l'observation de multiples données expérimentales sur le temps de réalisation d'opérateurs élémentaires

## Opérateurs

- K sélection de touche (*keystroke*)
- P pointage
- H repositionnement des mains sur un dispositif d'entrée
- D dessin d'une ligne à l'écran
- M action mentale
- R temps de réponse (éventuel) du système

# MODELES PREDICTIFS: KEYSTROKE



## Opérateurs: temps d'exécution moyens

- K appui simple touche 0,35"  
saisie secrétaire expérimenté(e) 0,22"  
saisie utilisateur moyen 0,28"  
saisie utilisateur novice 1,20"  
presser « Shift » ou « Ctrl » 0,08"
- P pointage souris vers une cible 1,10"  
clik souris 0,20"
- H repositionnement clavier 0,40"
- D dessin d'une ligne dépend de la longueur
- M action mentale 1,35"

$$T_{\text{exec}} = T_K + T_P + T_H + T_D + T_M (+ T_R)$$

# CONCLUSION











## Quel paradigme d'évaluation utiliser ?

		
<b>Expérimentation</b>	détection pbs	modification comportement
<b>Subjectif</b>	peu coûteux avis usager	précision
<b>Prédictif (modèles)</b>	systeme non nécessaire	portée limitée
<b>Analytique (experts)</b>	peu coûteux	grande expertise peut rater des pbs

**Nielsen** : expérience montrant qu'experts et sujets peuvent trouver des problèmes différents

# CONCLUSION

Quand utiliser un paradigme d'évaluation particulier ?

	conception	prototype	produit
<b>Études de terrain</b>			
<b>Évaluation analytique</b> (dont promenade cognitive, heuristique)			
<b>Expérimentation</b> (dont entretiens, questionnaires)			
<b><i>Quick and dirty</i></b>			

# CONCLUSION

## Evaluation *quick and dirty* : WWW

### Exemple : check-list d'évaluation rapide IBM (*IBM Web Design Guidelines*)

- L'objectif du site est-il clairement perçu ?
- Le public ciblé peut-il clairement s'identifier ?
- Le site est-il utile et pertinent pour ce public ?
- Le site est-il intéressant et attirant ? De quel point de vue ?
- Le site permet-il aux utilisateurs de réaliser toutes les tâches qu'ils veulent accomplir ?
- Les visiteurs peuvent-ils facilement accomplir ces tâches ?
- Le contenu et l'organisation du site sont-ils cohérents avec l'objectif ?
- Les informations importantes sont-elles faciles à trouver ?
- Les informations présentées sont-elles claires, faciles à comprendre ?
- Le visiteur sait-il toujours où il en est et comment faire pour aller où il veut ?
- Le graphisme est-il agréable ?

# CONCLUSION

- **L'utilisateur n'est pas tout ... et souvent n'est pas l'acheteur**
- **Typologie d'intéressées par le choix d'un logiciel**

[SESL: Ramage, 1997]

- Les utilisateurs du logiciel
- Leurs collègues et supérieurs (directeurs)
- Les développeurs et les revendeurs du logiciel
- Le service informatique de l'organisation (si nécessaire)
- Les clients de l'organisation
- Les syndicats et associations d'employés
- Les entreprises mères
- Les associations d'employés
- Les actionnaires
- Les gouvernements.

- **Mais l'utilisateur reste l'alpha et l'omega !**



# BIBLIOGRAPHIE

## Ouvrages de référence

Nielsen J. (1993) *Usability engineering*. Academic Press.

Baccino T. , Bellino C., Colombi T. (2005) *Mesure de l'utilisabilité des interfaces*. Hermès Sciences, Paris, France. ISBN 2-7462-1026-6

Stone D., Jarret C., Woodroffe M., Minocha S. (2005) *User Interface Design and Evaluation*. Morgan Kaufmann Publ. & Elsevier, San Francisco, CA. ISBN 0-12-088436-4.

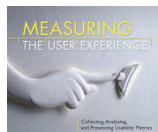
## Questionnaires accessibles sur le WWW

Gary Perlmann, *Web-Based User Interface Evaluation with Questionnaires* (non maintenu - consulté le 20/03/2014) : <http://oldwww.acm.org/perlman/question.html>

HCIL Team (U. Maryland), *QUIS : Questionnaire for User Interaction Satisfaction* (consulté le 20/03/2014) : [www.lap.umd.edu/QUIS/](http://www.lap.umd.edu/QUIS/)

## Forums ou sites WWW

**NN/g** Nielsen Norman group (rubrique User Testing) – *blog d'expert très actif sur l'évaluation*, (consulté le 20/03/2014) : [www.nngroup.com/topic/user-testing/](http://www.nngroup.com/topic/user-testing/)



Tom Tullis, Bill Albert website, *Measuring the User-Experience* – site assez riche recensant des articles sur l'évaluation, des pointeurs vers les outils etc... <http://www.measuringux.com/> (consulté le 20/03/2014)

# BIBLIOGRAPHIE

## Travaux cités

- Badiu R. (2014) Usability testing for mobile is easy. [Consulté en février 2014 sur : <http://www.nngroup.com/articles/mobile-usability-testing/>]
- Brooke J. (1996) SUS : a quick and dirty usability scale. In Jordan P.W. et al. (Eds). Usability Evaluation in Industry. Taylor & Francis, London.
- Chin J., Diehl V., Norman K. (1988) Development of an instrument measuring user satisfaction of the human-computer interface. Actes *ACM CHI'88*. 213-218.
- De Bono (1992) *Serious creativity*. Harper Business.
- Dumas J., Redish J. (1999) *A practical guide to usability testing*. Intellect, Exeter, UK.
- Hix D., Harston H.R. (1993) *Developing user interfaces : ensuring usability through product & process*. Wiley, New-York, NJ.
- Krueger R. (2000) *Focus groups: a practical guide for applied research*, Thousand Oaks, CA
- Lewis J. (1995) IBM computer usability satisfaction questionnaires : psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction*, 7 (1), 57-78.
- Kirakowski J., Corbett M. (1993) SUMI : the Software Usability Measurement Inventory. *British Journal of Educational Technology*, 24(3), 210-212.

# BIBLIOGRAPHIE

## Travaux cités (2)

- Nielsen J., Mack R., Bergendorf K., Grischkomsy N. (1986) Integrated software usage in the professional work environment : evidence from questionnaires and interviews. Actes *CHI'86*, New-York, NJ., ACM Press. 162-167.
- Nielsen J. Mach R. (1994) *Usability inspection methods*. John Wiley & Sons., New-York, NJ.
- Quesenbery W. (2003) Five dimensions of usability In. Albers et al. (Eds) *Content and complexity: information design in technical communication*. Lawrence Erlbaum. NJ.
- Tullis T; Stetson J. (2004) A comparison of questionnaires for accesing website usability. Proc. *UPA'2004*.
- Whiteside J., Bennet J., Holtzblatt K. (1988) Usability engineering: our experience and evolution. In Helander M. (Ed.) *Handbook of Human-Computer Interaction*. North-Holland, Amsterdam.