

LOG2420 — Analyse et conception d'interfaces utilisateur

Évaluation d'interfaces

Michel C. Desmarais

Génie informatique et logiciel
École Polytechnique de Montréal

18 août, 2021

Types d'évaluation

Principaux types :

- ▶ **Tests d'utilisabilité** : approche empirique
- ▶ **Évaluations heuristiques** : approche experte
- ▶ Entrevues et questionnaires de satisfaction
- ▶ Inspection cognitives

Les approches sont complémentaires.

- ▶ Les tests d'utilisabilité et évaluations expertes servent en cours et en fin de développement.
- ▶ Les entrevues et questionnaires de satisfaction peuvent couvrir les aspects qualitatifs après un test d'utilisabilité ou suite à un déploiement.
- ▶ Les inspections cognitives servent surtout en phase de conception.

Le principe de l'évaluation

- ▶ Même principes que les tests logiciels
 - ▶ Toute exigence d'utilisabilité devrait faire l'objet d'un test d'utilisabilité empirique
 - ▶ Les exigences d'utilisabilité doivent être vérifiables
- ▶ L'évaluation d'une interface doit donc se faire par rapport à des **tâches**, des **utilisateurs** et des **objectifs d'utilisabilité** spécifiques.

Tests d'utilisabilité

Méthode

- ▶ Utilisateurs représentatifs
- ▶ Tâches représentatives et exhaustives
- ▶ Analyse du protocole de pensée à voix haute
- ▶ Différents outils s'offrent pour faciliter la tâche d'analyse :
 - ▶ Observation par plusieurs ergonomes
 - ▶ Enregistrement du “click through” et “playback”
 - ▶ Vidéo
 - ▶ Miroir à sens unique

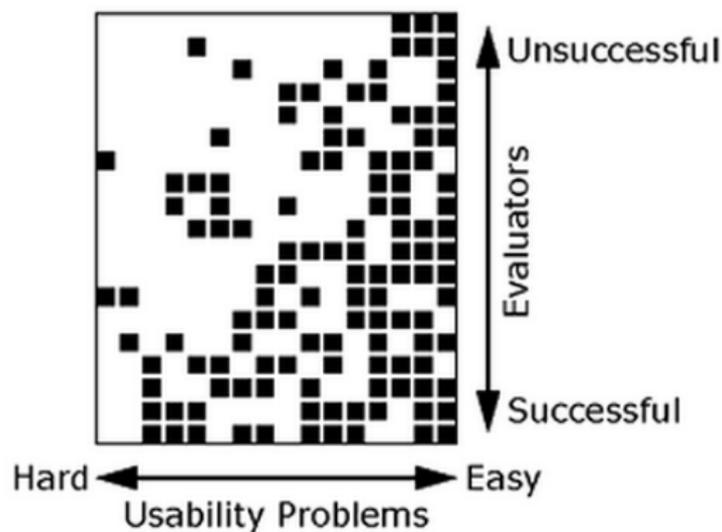
Définition des tâches

- ▶ Comment déterminer les bonnes tâches ?
 - ▶ La priorité dépend de :
 - ▶ Fréquence : les plus fréquentes sont plus prioritaires
 - ▶ Criticalité : les plus incontournables
 - ▶ Représentativité : correspond aux besoins des utilisateurs
- ▶ Granularité :
 - ▶ Niveau adéquat : correspond à des buts relativement autonomes
- ▶ Vérifiables :
 - ▶ Le succès peut être validé

Combien d'experts ou d'évaluateurs ?

Expérience de Nielsen et Landauer (1993)

<http://www.nngroup.com/articles/how-to-conduct-a-heuristic-evaluation/>



Combien d'utilisateurs/experts pour l'évaluation ?

- ▶ Le nombre de problèmes restants suit la courbe suivante¹ :

- ▶ $N_0(1 - (1 - L)^k)$

- ▶ Où :

N_0 : nombre de problèmes au total

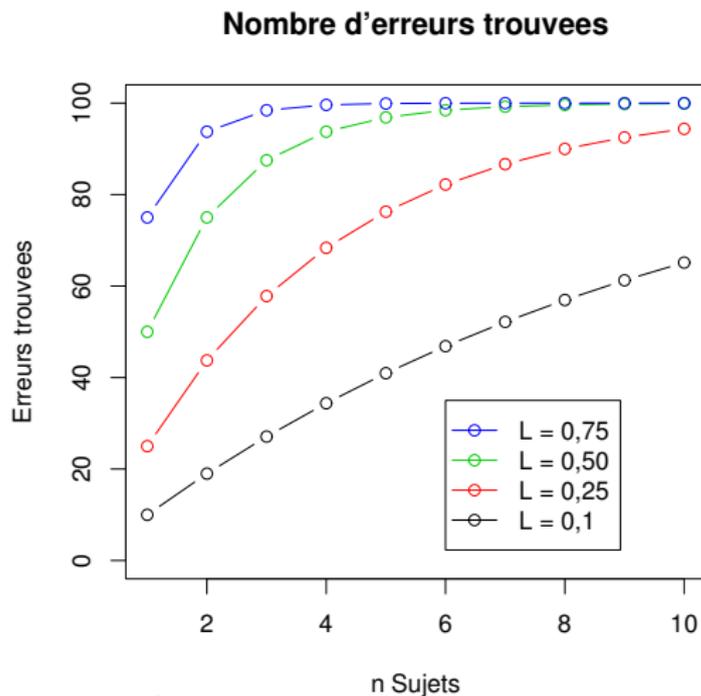
L : proportion de problèmes identifiés par un sujet

k : nombre de sujets

1. Nielsen, Jakob, and Landauer, Thomas K. : "A mathematical model of the finding of usability problems," Proceedings of ACM INTERCHI'93 Conference (Amsterdam, The Netherlands, 24-29 April 1993), pp. 206-213.

Relations entre le nombre de problèmes trouvés et le nombre de sujets

$$N_0(1 - L)^k$$



Solution pour estimer le nombre de problèmes

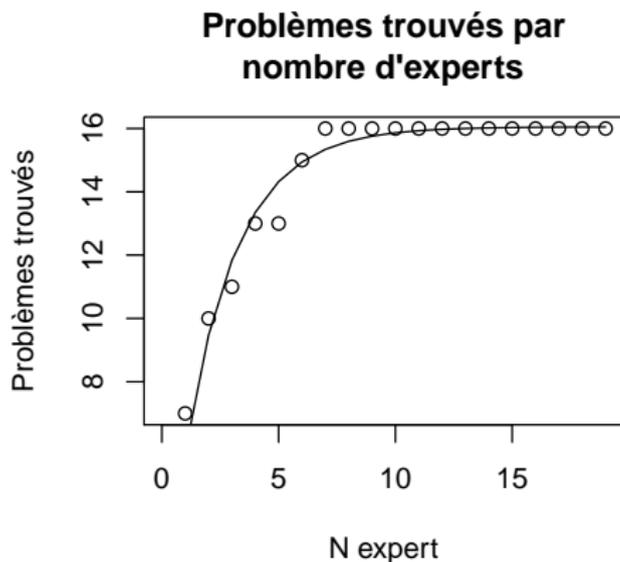
- ▶ Le problème est de la même famille que la loi de la décroissance exponentielle :
 - ▶ $N_0 e^{-\lambda t}$
 - ▶ Où :
 - N_0 : nombre de problèmes initiaux
 - λ : taux de décroissance (des problèmes)
 - t : temps sur une échelle continue (nombre de sujets pour nous)
- ▶ En spécifiant le nombre de problèmes restants à l'itération t comme $N_t = N_0(1 - L)^t$, et en posant $e^{-\lambda} = (1 - L)$, le problème s'exprime sous la forme de la loi de décroissance :
 - ▶ $N_t = N_0 e^{-t\lambda}$
- ▶ Cette loi correspond à un modèle d'évolution du nombre de bogues restants suite à des itérations de tests logiciels^{2 3}

2. Goel, A.L., and Okumoto, K. (1979). A time dependent error detection model for software reliability and other performance measures. *IEEE Trans. Reliability*, vol R-28, 1979, 206–211.

3. Wood, A. (1996) "Predicting software reliability." *Computer*, vol.29, no.11, 69–77.

Exemple avec données de Nielsen et Landauer

- ▶ Un exemple de calculs du nombre de problèmes en fonction des essais est fourni au lien <http://www.groupe.polymtl.ca/log2420/Seances/Evaluation/courbe-exp.html>
- ▶ Données tirées de Nielsen et Landauer



Laboratoire d'utilisabilité

- ▶ Un miroir à sens unique
- ▶ Un écran qui réplique celui de l'utilisateur (aussi enregistré sur vidéo)
- ▶ Un vidéo qui filme l'utilisateur
- ▶ Des outils de montage de vidéo
- ▶ Image (Sun Microsystem)
- ▶ Sondage, Nielsen, 1994 :
<http://www.useit.com/papers/uselabs.html>

Métriques

Utilisation des métriques

- ▶ Établir un barème minimal à respecter pour des spécifications
- ▶ Identifier et quantifier des éléments problématiques de l'interface
- ▶ Comparer différentes versions
- ▶ Comparer entre produits ou concepts concurrents
- ▶ Utilisation de métriques dans différents contextes :
 - ▶ Un élément de performance pour l'évaluation du succès (bonis)
 - ▶ Élément pour entente contractuelle

Métriques objectives

- ▶ Pourcentage de tâches complétées en un temps donné
- ▶ Pourcentage de tâches réussies
- ▶ Temps de complétion de tâches
- ▶ Nombre d'erreurs
- ▶ Temps de récupération des erreurs
- ▶ Courbe d'apprentissage

Métriques subjectives

- ▶ Cinq catégories
 - ▶ (1) efficacité, (2) affectif, (3) assistance, (4) contrôle, (5) facilité d'apprentissage
- ▶ Mesurable par questionnaire, par exemple :
 - ▶ Ce logiciel répond trop lentement
 - ▶ Je recommanderais ce logiciel à mes collègues
 - ▶ Les instructions sont utiles et accessibles
 - ▶ Je me demande parfois si j'utilise la bonne commande
 - ▶ Travailler avec ce logiciel est efficace

Questionnaires

- ▶ Échelles d'évaluation d'attitude, de satisfaction ou d'appréciation
- ▶ De petits détails dans la formulation d'une question peuvent faire une énorme différence
 - ▶ Exemple : "votre écran est-il monochrome?"
- ▶ Les questionnaires sont comme des interfaces :
 - ▶ Ils doivent être testés !

Conseils pratiques

- ▶ Importance des vidéos
 - ▶ Souvent le moyen le plus convaincant
 - ▶ Ex. de Xerox PARC
(<https://www.youtube.com/watch?v=DUwXN01ARYg>)
- ▶ Démontrer l'impact en termes financiers ou autre
- ▶ Opter pour les tests plus souvent possible

Difficultés méthodologiques

- ▶ L'effet Hawthorne
 - ▶ Étude dans l'Illinois du travail dans l'usine Hawthorne, 1939
 - ▶ Effet de l'éclairage sur le travail
- ▶ Dissonance cognitive
 - ▶ L'effet de la rémunération sur l'appréciation de l'expérience
- ▶ L'Effet de halo
- ▶ L'effet de proximité dans le temps (recency effect)
- ▶ L'effet d'ordre des tâches/questions (mesures répétées)
- ▶ Les effets de groupe
- ▶ ...et les mille et un dédales de l'expérimentation avec des sujets humains

Outil intéressant

- ▶ Questionnaire Web de Perlman [http:](http://www.acm.org/~perlman/question.cgi?form=CSUQ)

`//www.acm.org/~perlman/question.cgi?form=CSUQ`

Évaluations heuristiques

Méthode d'évaluations heuristiques

- ▶ Nécessite une maîtrise des règles heuristiques
 - ▶ Nielsen
 - ▶ Bastien et Scapin
- ▶ Dure en général d'une à deux heures
- ▶ Utile d'assister l'évaluateur par un aide
- ▶ On fournit une liste de tâches ou/et des scénarios d'utilisation
- ▶ Les problèmes identifiés sont ensuite classifiés selon leur gravité

Efficacité et efficience des tests heuristiques

- ▶ Selon plusieurs, la méthode la plus rentable
- ▶ Sujet à la même loi quant au nombre de problèmes identifiés par utilisateur
 - ▶ Toutefois, le 'L' peut être fort différent des tests utilisateurs
 - ▶ Exemple d'évaluation avec 19 évaluateurs et 16 problèmes d'utilisabilité
(http://www.useit.com/papers/heuristic/heuristic_evaluation.html)