

# IFT 2905: INTERFACES PERSONNE-MACHINE

Évaluation: **tests d'utilisabilité**

<http://tiny.cc/ift2905>



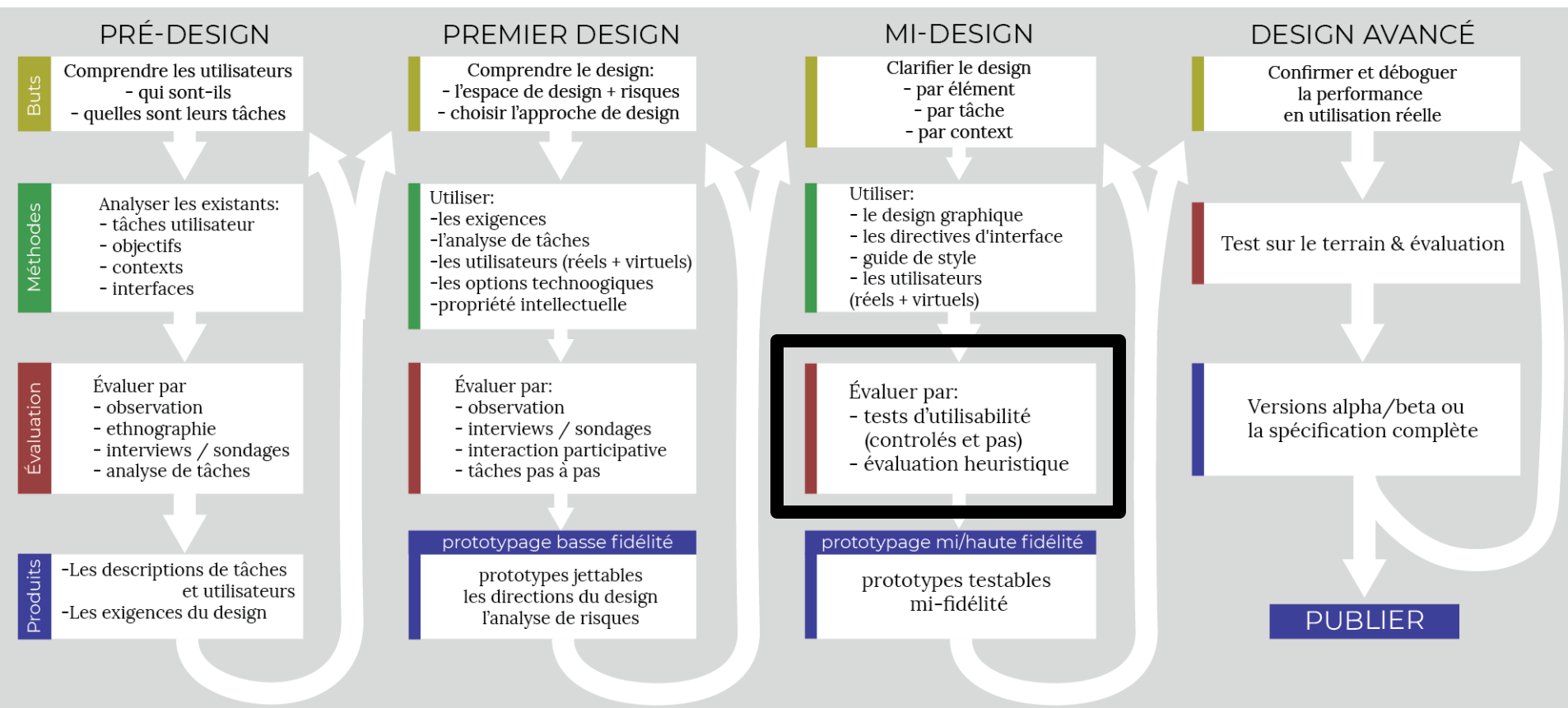
Image © Katerina Kamprani, [www.theuncomfortable.com](http://www.theuncomfortable.com)

## Mikhail Bessmeltsev

Certaines diapos de Yoon, McLean et Thomaszewski

Rappel

# DESIGN CONCEPTUEL



*Rappel*

	<i>formative</i>	<i>sommative</i>
genre d'apprentissage	exploration	évaluation
type de données	qualitatives	quantitative
niveau de contrôle	moins	plus
formalité	plutôt moins	plutôt plus
phase	pré/premier/mi-design	mi-design/design avancé
coût	souvent moins cher	souvent cher
tâches utilisateur	relativement ouvertes	attribuées

Dérivé d'un diapo de "Evaluating User Interfaces"  
par University of Minnesota

# TESTS D'UTILISABILITÉ

= études d'utilisabilité

évaluation **sommative** qui détermine si  
l'interface peut être utilisée par *les*  
*utilisateurs directs* pour faire les tâches *pour*  
*lesquelles elle a été conçue*

# TESTS D'UTILISABILITÉ

nécessitent un prototype raffiné

- nouveau
- ou existant

Comparez avec les évaluations avant pour mieux comprendre les utilisateurs, les tâches et les exigences

*Rappel*

	<i>formative</i>	<b>sommative</b>
genre d'apprentissage	exploration	évaluation
type de données	qualitatives	quantitative
niveau de contrôle	moins	plus
formalité	plutôt moins	plutôt plus
phase	pré/premier/mi-design	mi-design/design avancé
coût	souvent moins cher	souvent cher
tâches utilisateur	relativement ouvertes	attribuées

Dérivé d'un diapo de "Evaluating User Interfaces"  
par University of Minnesota

# **BUTS DE TESTS D'UTILISABILITÉ**

dépendent de vos buts d'évaluation, mais normalement on mesure:

- performance
- erreurs
- problèmes dans l'apprentissage
- acceptance

# QUANTIFICATION DE L'UTILISABILITÉ

Utilisabilité  $\approx f(\textit{efficacité}, \textit{efficacité}, \textit{satisfaction})$

## Dimensions de l'utilisabilité

**Facilité d'apprentissage:** est-ce facile à apprendre?

**Mémorabilité:** est-ce facile de se rappeler le fonctionnement?

**Erreurs:** les erreurs sont-elles peu nombreuses et récupérables?

...

But: quantifier ces mesures d'utilisabilité  
par des expériences



# MÉTHODES

On peut utiliser presque toutes les méthodes qu'on connaît:

- observation
- interview
- sondage/questionnaire
- analyse de journaux/logs

# COMMENT CRÉER UN TEST D'UTILISABILITÉ

- Méthodes: triangulation
  - P.e. une pour **compter** (*quantitative*) une pour **interpréter** (*qualitative*)
- Métriques: dépendent des objectifs de l'évaluation
  - Efficacité
    - Taux d'achèvement =  $\frac{\# \text{ de tâches réussies}}{\# \text{ total de tâches essayées}}$
    - Nombre d'erreurs
  - Efficience
    - Efficience basée sur le temps (**Time-based Efficiency**)
    - Efficience relative globale (**Overall relative Efficiency**)
  - Satisfaction
    - ASQ, NASA-TLX, SMEQ, UME, SEQ



# EXEMPLES DE MÉTRIQUES

## temps pour

- une tâche complète
- une sous-tâche (p.e. trouver quelque chose sur l'écran)
- apprendre comment faire une tâche
- reprendre une tâche après une interruption
- atteindre un niveau de compétence

## erreurs

- # par tâche, par unité de temps, etc.
- # d'utilisateurs qui font l'erreur
- etc.

# EXEMPLES DE MÉTRIQUES

## événements d'intérêt

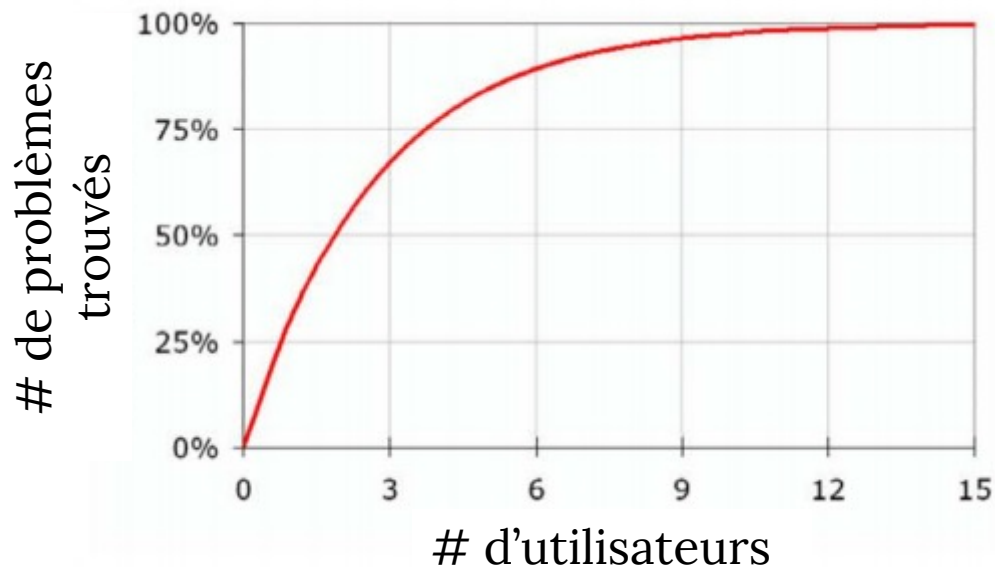
- pages vues ou clics
- accès à des outils particuliers
- délais d'attente
- questions posées ou outils d'aide consultés
- # utilisateurs prêts à recommander

## facteurs subjectifs

- facilité d'utilisation
- charge de travail (NASA TLX)
- satisfaction
- esthétique
- préférence
- ...

# COMBIEN D'UTILISATEURS?

5-12 est considéré comme un nombre acceptable



doit être représentatif de vos groupes d'utilisateurs  
p.e. si vous voulez soutenir les utilisateurs experts et  
novices, il faut avoir un bon nombre des deux

# TÂCHES À FAIRE

Normalement, l'expérimentateur spécifie la tâche

Utilisez les descriptions de tâches!  
(sauf le contexte, l'histoire, etc.)

- tâche entière
- sous-tâche/tâche concrète
  - p.e. la sous-tâche qui vous mènera d'un écran à l'autre

# VIDÉO

<https://youtu.be/3Qg80qTfzgU>

en anglais, sous-titres traduits  
automatiquement



# VIDÉO ADDITIONNELS

Comment exécuter un test d'utilisabilité à distance

<https://youtu.be/ZkDafFDtJ1Y>

Zipcar

<https://youtu.be/QckIzHC99Xc?t=56s>

# AUTRE MÉTHODES DE TEST D'UTILISABILITÉ

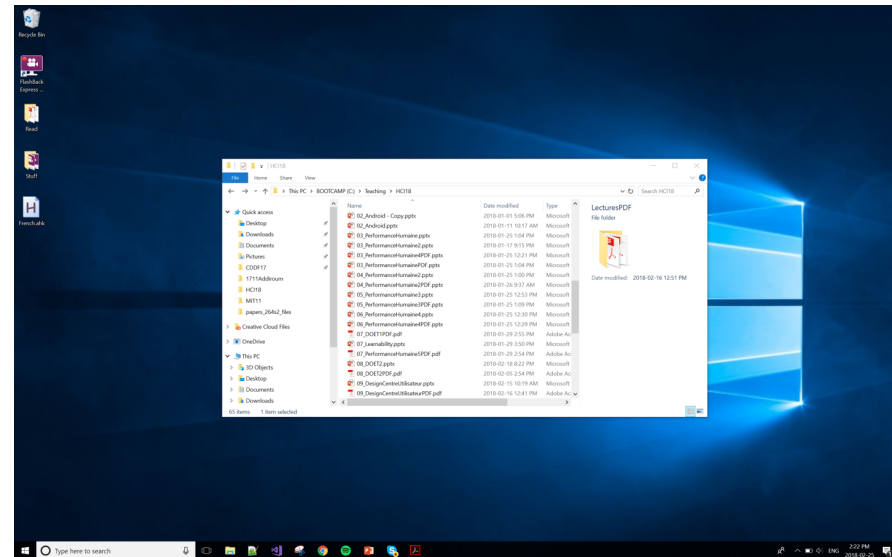
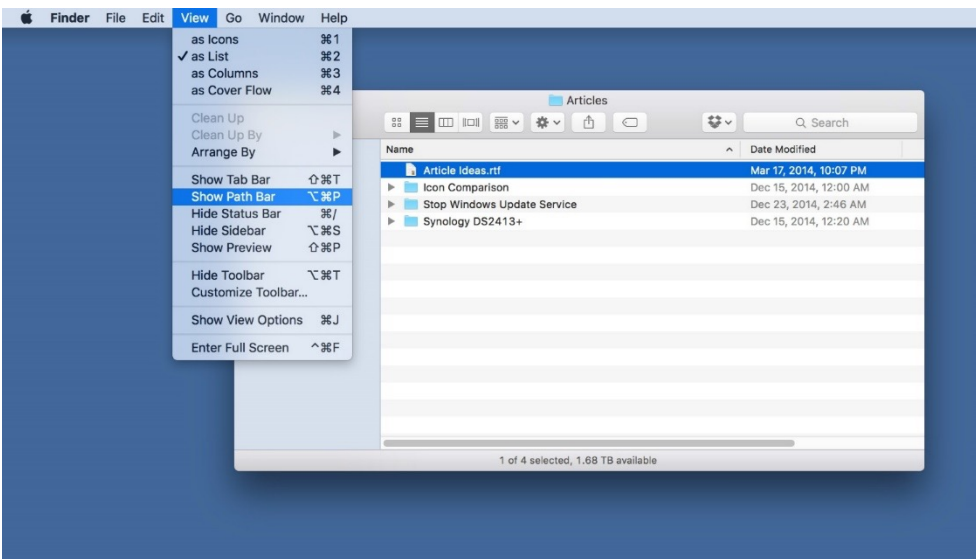
Sans utilisateur (moins cher!)

- *Cognitive walkthrough*
- Évaluation heuristique

# EXPÉRIENCE CONTRÔLÉE

Commencer avec une hypothèse testable  
(quantifiable et mesurable)

*p.e.*, la barre de menu Mac est plus rapide que celle de Windows



# EXPÉRIENCE CONTRÔLÉE

- Commencer avec une hypothèse testable (quantifiable et mesurable)
- Manipuler des variables indépendantes
- Mesurer les variables dépendantes
- Utiliser des méthodes statistiques pour analyser

# EXPÉRIENCE CONTRÔLÉE

- Commencer avec une hypothèse testable (quantifiable et mesurable)
  - «la barre de menu Mac est plus rapide que celle de Windows»
- Manipuler des variables indépendantes
- Mesurer les variables dépendantes
- Utiliser des méthodes statistiques pour analyser

# EXPÉRIENCE CONTRÔLÉE

- Commencer avec une hypothèse testable (quantifiable et mesurable)
  - «la barre de menu Mac est plus rapide que celle de Windows»
- Manipuler des variables indépendantes
  - différentes interfaces, classes d'utilisateurs, tâches dans l'exemple, la position  $y$  de la barre de menu
- Mesurer les variables dépendantes
- Utiliser des méthodes statistiques pour analyser

# EXPÉRIENCE CONTRÔLÉE

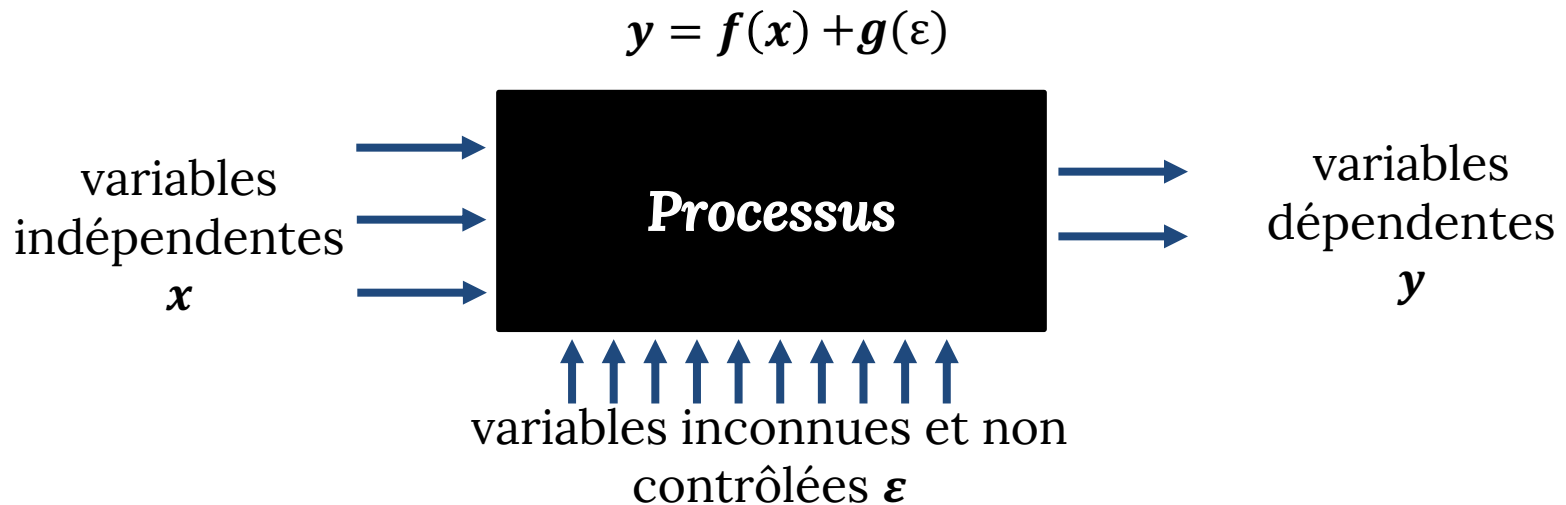
- Commencer avec une hypothèse testable (quantifiable et mesurable)
  - «la barre de menu Mac est plus rapide que celle de Windows»
- Manipuler des variables indépendantes
  - différentes interfaces, classes d'utilisateurs, tâches dans l'exemple, la position  $y$  de la barre de menu
- Mesurer les variables dépendantes
  - temps, erreurs, numéro de tâches accomplies, satisfaction
- Utiliser des méthodes statistiques pour analyser

# EXPÉRIENCE CONTRÔLÉE

- Commencer avec une hypothèse testable (quantifiable et mesurable)
  - «la barre de menu Mac est plus rapide que celle de Windows»
- Manipuler des variables indépendantes
  - différentes interfaces, classes d'utilisateurs, tâches dans l'exemple, la position  $y$  de la barre de menu
- Mesurer les variables dépendantes
  - temps, erreurs, numéro de tâches accomplies, satisfaction
- Utiliser des méthodes statistiques pour analyser
  - comment les changements dans les variables indépendantes affectent les variables dépendantes
  - si ces effet sont signifiants (***vraie causalité ou du bruit?***)
  - si on devrait accepter ou rejeter l'hypothèse



# CONCEPTION D'EXPÉRIENCE



- Processus de boîte noire (***p.e., accès à la barre de menu***)
- ***f*** est la fonction qui décrit la relation entre  $x$  et  $y$  (***qu'on veut trouver***)
- ***g*** est la fonction qui décrit la relation entre  $\varepsilon$  et  $y$  (***dont on cherche à minimiser l'impact***)

But: concevoir des expériences telle que l'influence des variables non contrôlées soit négligeable

# EXEMPLE MENUBAR

Les variables non contrôlées pour l'exemple menubar incluent

- le dispositif de pointage utilisé (souris, track ball, écran tactile, ...)
- où le pointeur de la souris a commencé
- la surface sur laquelle ils déplacent la souris
- le niveau de fatigue de l'utilisateur
- leur expérience passée avec une sorte de menubar ou l'autre

Toutes ces variables sont inconnues et pourraient affecter la variable dépendante (vitesse d'accès)

# CONCEPTION DE L'EXPÉRIENCE MENUBAR I

## Les utilisateurs

Utilisateurs Windows ou utilisateurs Mac?

Droitier ou gaucher?

Comment choisir (échantillonnage)?

## La mise en œuvre

Windows réel et Mac réel

Gestionnaire de fenêtres artificiel qui nous permet de contrôler la position de la barre de menu

## Les tâches

Réaliste: traitement de texte, courriel, navigation sur le Web

Artificiel: pointant à plusieurs reprises sur la fausse barre de menu

# CONCEPTION DE L'EXPÉRIENCE MENUBAR II

## Mesurer le temps

- quand commence et finit le mouvement?

## Ordre des tâches et des conditions d'interface

- risque d'apprentissage
- risque de fatigue

## Matériel

- souris, trackball, écran tactile, joystick?
- PC ou Mac? Quelle machine particulière?

# CONCEPTION D'EXPÉRIENCE – PRÉOCCUPATIONS

## Validité interne

- Les résultats observés sont-ils réellement **causés** par les variables indépendantes?
- Confondement: une variable qu'on ne contrôlait pas a un effet systématique sur les variables dépendantes

## Validité externe

Les résultats observés peuvent-ils être **généralisés** au monde extérieur au laboratoire?

## Fiabilité

Des résultats cohérents seront-ils obtenus en **répétant** l'expérience

# EXEMPLE: COMBIEN DE BILLES DANS CHAQUE BOÎTE?

- Hypothèse: la case **A** a un nombre différent de billes que la case **B**



## Fiabilité

Le comptage manuel des billes n'est fiable que s'il y a peu de billes  
Le comptage répété améliore la fiabilité (méthodes statistiques)

## Validité interne

Supposons que nous pesons les boîtes au lieu de compter les billes  
Que faire si une bille **A** a un poids différent d'une bille **B**?  
Que faire si les boîtes ont elles-mêmes des poids différents?

## Validité externe

Ce résultat s'applique-t-il à toutes les boîtes **A** et **B**?

# MENACES À LA VALIDITÉ INTERNE I

## Deux stratégies pour la validité interne

- randomisation: empêche les variables inconnues d'avoir des effets systématiques sur les variables dépendantes
- contrôle: essaie de maintenir des variables inconnues constantes

## Effets de l'ordre

- Problème: Les gens apprennent et les gens se fatiguent
- Solution: Ne pas présenter les tâches ou les interfaces dans le même ordre pour tous les utilisateurs (randomiser ou équilibrer)

## Effets de sélection

- Problème: Sélectionner des groupes peut créer du biais
- Ne pas utiliser de groupes préexistants (sauf si le groupe est une variable indépendante)
- Attribuer aléatoirement des utilisateurs à des variables indépendantes

# BIAIS DU FACILITATEUR

Problème: Le facilitateur peut être enthousiaste à propos de l'interface  $X$  mais pas  $Y$

## Solutions

- Donner de la formation et des séances d'information sur papier, pas en personne
- Les expériences en **double aveugle** empêchent le sujet et le facilitateur de savoir si c'est la condition  $X$  ou  $Y$  (p.e., médicaments)
- → essentiel si la mesure des variables dépendantes nécessite du jugement subjective



# MENACES À LA VALIDITÉ EXTERNE

*Les résultats s'appliquent-ils au-delà du laboratoire?*

Population: les échantillons sont-ils représentatifs?

Choisir des échantillons aléatoires de la population cible

Écologique: les conditions sont-elles comme le monde réel?

Rendre les conditions de laboratoire aussi réalistes que possible dans des aspects importants

Formation: la formation est-elle réaliste?

La formation devrait imiter comment une interface réelle serait rencontrée et apprise

Tâche: les tâches sont-elles sont réalistes et représentatives?

Baser les tâches sur l'analyse des tâches

# MENACES À LA FIABILITÉ - VARIATION INCONTRÔLÉE

## Expériences précédentes

Séparer novices et experts en différentes classes, ou utiliser une seule classe?

## Différences entre utilisateurs

Les utilisateurs les plus rapides sont 10 fois plus rapides que les utilisateurs les plus lents

## Erreur de mesure

Le temps passé à la tâche peut inclure tousser et autres distractions

# MENACES À LA FIABILITÉ – SOLUTIONS

Éliminer la variation systématique en maintenant constantes les variables incontrôlées

- Sélectionner les utilisateurs pour certaines expériences qu'ils ont (ou n'ont pas)
- Donner à tous les utilisateurs la même formation

Éliminer la variation aléatoire (*le bruit*)

- Répétition: beaucoup d'utilisateurs, de nombreux essais
- Plus la **variance** est grande, plus d'échantillons sont nécessaires pour avoir une **moyenne** fiable

# DIVISION EN BLOCS

- Diviser les utilisateurs en sous-ensembles plus homogènes que l'ensemble
  - Exemple: test du taux d'usure de différentes semelles de chaussures
  - Beaucoup de variation entre les pieds de différents enfants
  - Mais les pieds sur le même enfant sont beaucoup plus homogènes

→ Chaque enfant est un bloc
- Appliquer toutes les conditions dans chaque bloc
  - Mettre le matériau **A** sur un pied, le matériau **B** sur l'autre
- Mesurer la différence dans chaque bloc
  - $Usure(\mathbf{A}) - Usure(\mathbf{B})$
- Randomiser dans le bloc pour éliminer les menaces de validité interne
  - Mettre au hasard **A** sur le pied gauche ou le pied droit

# DESIGN INTER-SUJETS

*Une conception d'expérience standard sans division en blocs*

Les utilisateurs sont aléatoirement divisés en deux groupes

- Un groupe ne voit qu'une interface

Les résultats sont comparés **entre** différents groupes

Est-ce que la moyenne( $x_i$ ) > moyenne( $y_i$ )?

Élimine la variation due aux effets d'ordre

Les utilisateurs ne peuvent pas apprendre d'une interface pour faire mieux sur l'autre

Peut avoir des problèmes de fiabilité

Liés aux différences entre les deux groupes d'utilisateurs

# DESIGN INTRA-SUJETS

*Une conception d'expérience standard avec division en blocs*

Chaque utilisateur voit à la fois l'interface  $X$  et  $Y$   
(dans un ordre aléatoire)

Les résultats sont comparés pour chaque  
utilisateur

Pour l'utilisateur  $i$ , on calcule la différence  $x_i - y_i$

Est-ce que la moyenne( $x_i - y_i$ )  $> 0$ ?

Élimine la variation due aux différences de l'utilisateur

Les utilisateurs ne se comparent qu'à eux-mêmes

Peut avoir des problèmes de validité interne dus  
aux effets de l'ordre

# EQUILIBRAGE (COUNTERBALANCING)

Éliminer les effets d'ordre en changeant systématiquement l'ordre des conditions

Randomisation n'est pas efficace si le nombre d'utilisateurs est petit

Carré latin (*latin square* → *Sudoku*)

- attribuer aléatoirement des sujets à des groupes de taille égale
- Pour **N** conditions (A, B, C, ...) on fait **N** groupes
- Le carré latin garantit que chaque condition se produit dans chaque position

G1	G2
A	B
B	A

2x2

G1	G2	G3
A	C	B
B	A	C
C	B	A

3x3

G1	G2	G3	G4
A	D	C	B
B	A	D	C
C	B	A	D
D	C	B	A

4x4

... etc.

# PIAZZA VS STUDIUM

The screenshot shows the Piazza LMS interface for a class named 'IFT 6113'. The top navigation bar includes 'Q & A', 'Resources', 'Statistics', and 'Manage Class'. The main content area is titled 'Class at a Glance' and provides a summary of the class's activity:

- 1 unread posts**
- no unanswered questions**
- no unresolved followups**

Additional statistics shown include:

- license status:** license not needed
- 141 total posts**
- 656 total contributions**
- 41 instructors' responses**
- 10 students' responses**
- 206 min avg. response time**

The 'Student Enrollment' section shows 12 enrolled students out of 5 estimated. Below this, there is a 'Share Your Class' section with instructions for professors and a link to download the app from the App Store or Google Play.

The left sidebar contains a 'PINNED' section with posts such as 'Paper presentation: Thursda...' and 'Important links', and a 'THIS WEEK' section with 'Text Editor Improvements'.

The screenshot shows the Studium LMS interface for a forum titled 'IFT2905-A-H21 - Interfaces personne-machine'. The forum is categorized under 'Mes cours / IFT2905-A-H21 / Introduction / Questions'. A search bar and a link to 'Administration de l'activité' are visible.

The forum content includes a section for 'Questions' with the following text: 'Il sera possible de poser des questions dans ce forum par rapport au cours et aux devoirs.' Below this is a button to 'Ajouter une discussion'.

The forum posts are listed in a table with the following columns: Discussion, lancée par, Dernier message, Réponses, and M'abonner.

Discussion	lancée par	Dernier message	Réponses	M'abonner
☆ Numéro 2.2 du devoir 4	Thalée St-Jacqu... 31 mars 2021	Thalée St-Jacqu... 31 mars 2021	2	<input type="checkbox"/>
☆ Devoir 4 - Premier Design question 2	William Bach 28 mars 2021	Steve Levesque 29 mars 2021	1	<input type="checkbox"/>
☆ Faire l'implémentation en HTML/CSS/JS au lieu d'Android.	Nassim Sabou... 25 mars 2021	Mikhail Bessm... 25 mars 2021	1	<input type="checkbox"/>
☆ Detail du fonctionnement de notre application	Jean-Marc Pru... 14 mars 2021	Mikhail Bessm... 14 mars 2021	2	<input type="checkbox"/>