

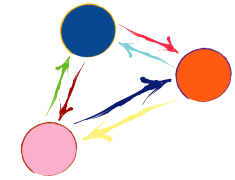
Interaction Homme-Machine

L'évaluation, phase si difficile mais cruciale !

Gaëlle Calvary
Professeur en Informatique

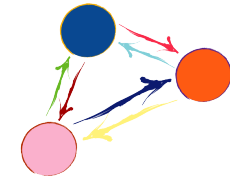
Institut polytechnique de Grenoble
Laboratoire d'Informatique de Grenoble



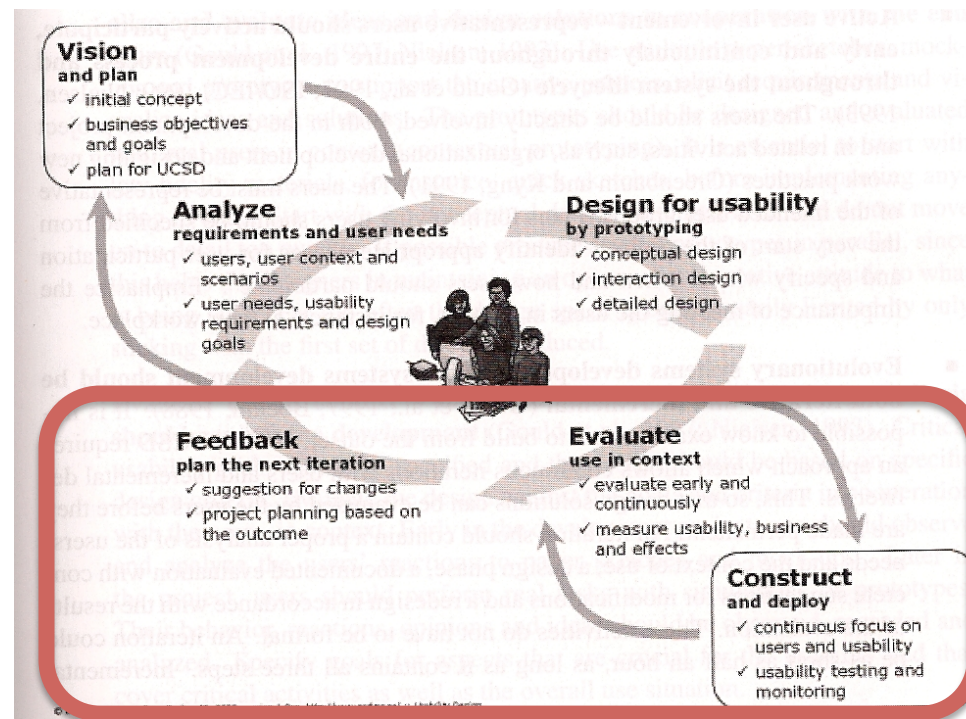


Principes généraux

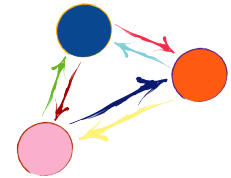
Principes généraux



#1. L'évaluation est fondamentale ! Les prototypes ne sont qu'un moyen d'évaluation !



Principes généraux

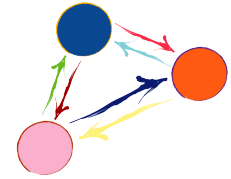


#1. L'évaluation est fondamentale !

#2. Vaste arsenal de techniques et d'outils complémentaires. Deux grandes classes d'approches :

- prédictives : sans utilisateur
- expérimentales : avec utilisateurs

Principes généraux



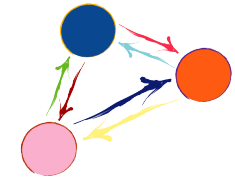
#1. L'évaluation est fondamentale !

#2. Vaste arsenal de techniques et d'outils complémentaires

#3. Evaluer au plus tôt ! Deux finalités complémentaires :

- évaluation : formative
- validation : sommative

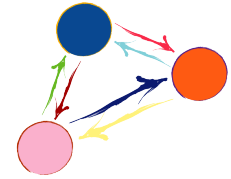
Principes généraux



- #1. L'évaluation est fondamentale !
- #2. Vaste arsenal de techniques et d'outils complémentaires
- #3. Evaluer au plus tôt !
- #4. Savoir ce que l'on cherche ! Rien de plus frustrant que d'apprendre par 30 utilisateurs que « le bouton est trop petit » !! Ces problèmes d'utilisabilité font écran par rapport aux retours attendus quant à la valeur (worth) du système !!!

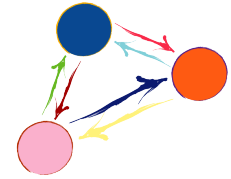
« Don't waste users on the small stuff. Critique can identify minor issues that can be resolved before testing, allowing users to focus on the big issues » (S. Klemmer)

Principes généraux



- #1. L'évaluation est fondamentale !
- #2. Vaste arsenal de techniques et d'outils complémentaires
- #3. Evaluer au plus tôt !
- #4. Savoir ce que l'on cherche !
- #5. Concevoir l'évaluation ! L'évaluation se travaille, s'anticipe, se planifie, se chiffre. L'évaluation a un coût ; la non évaluation a un surcoût !!!

Principes généraux



#1. L'évaluation est fondamentale !

#2. Vaste arsenal de techniques et d'outils complémentaires

#3. Evaluer au plus tôt !

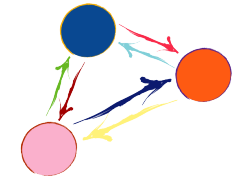
#4. Savoir ce que l'on cherche !

#5. Concevoir l'évaluation !

#6. Mieux vaut une évaluation mal conçue que pas d'évaluation du tout !

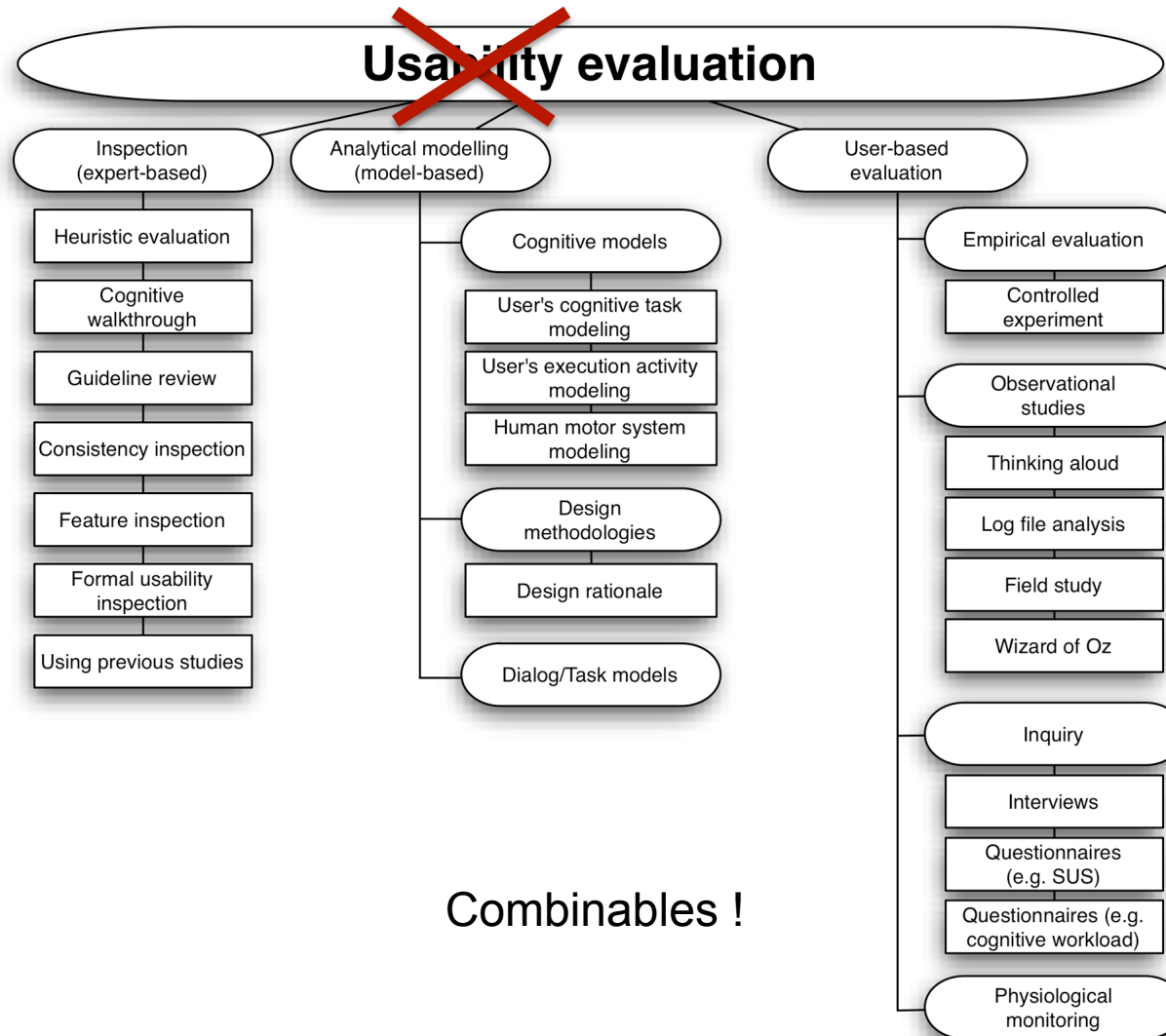
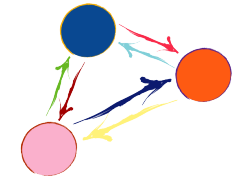
Une évaluation est toujours riche d'enseignements !

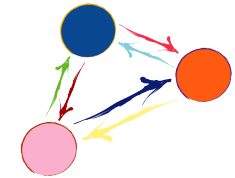
... mais mieux vaut une évaluation bien conçue !



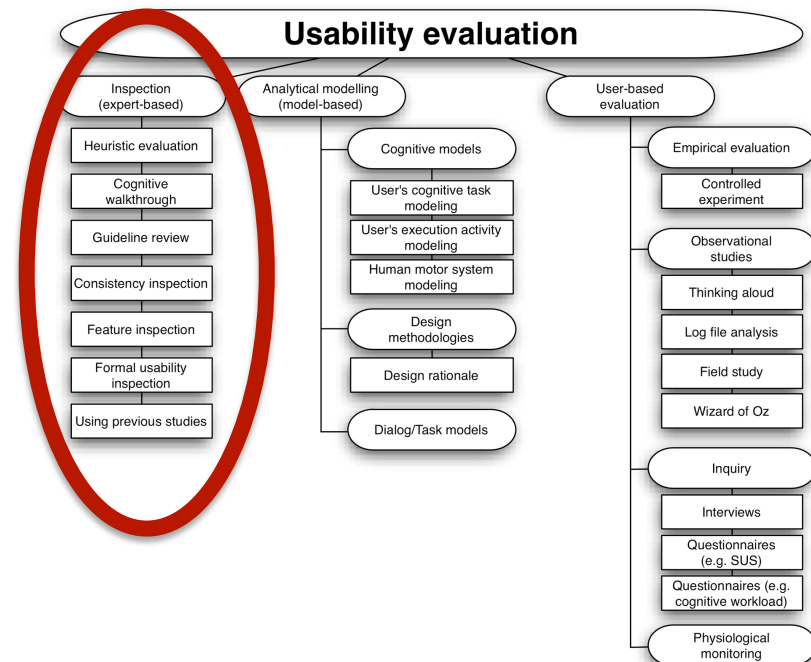
Panorama général des approches

Classification des approches

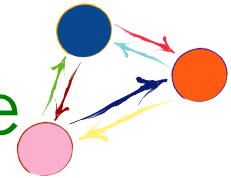




Inspection ergonomique

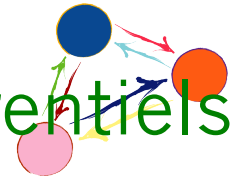


Evaluation heuristique : évaluation experte



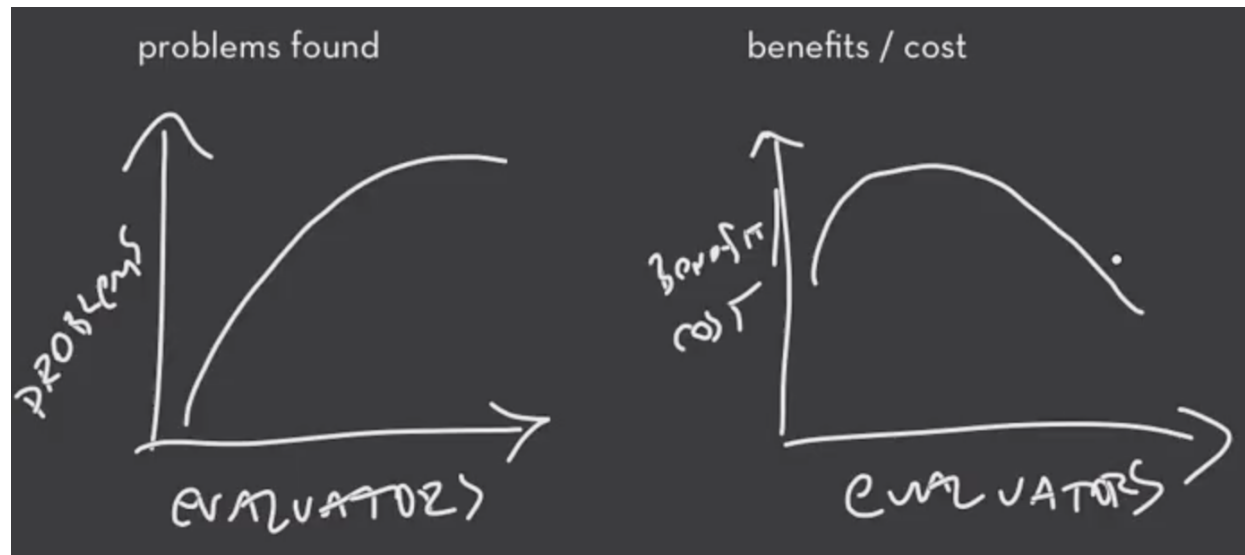
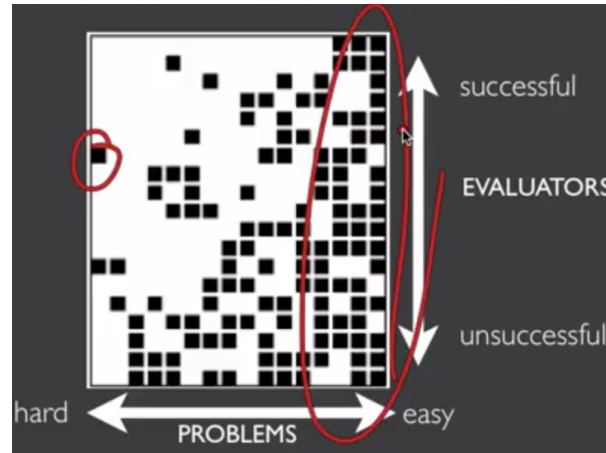
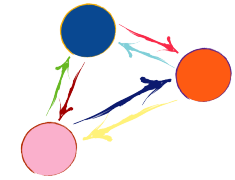
- Principe : examen de l'IHM sur la base de référentiels d'utilisabilité
- De nombreux promoteurs et de nombreux référentiels : Jakob Nielsen, Ben Schneidermann, ..., D. Scapin et C. Bastien
- Ces référentiels servent aussi à appuyer les choix de conception (cf cours précédent)
- En pratique :
 - 3 à 5 évaluateurs inspectent l'IHM puis confrontent leurs avis
 - S'applique à tout niveau de fidélité

Evaluation heuristique : exemples de référentiels



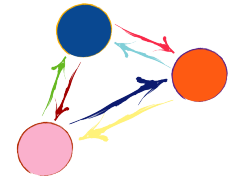
- D. Scapin et C. Bastien
 - Compatibilité
 - Guidage
 - Charge de travail
 - Homogénéité-Cohérence
 - Gestion des erreurs
 - Contrôle explicite
 - Adaptabilité
 - Signifiante des codes et dénominations
- J. Nielsen
 - Visibility of System Status
 - Match between System & World
 - User Control & Freedom
 - Consistency & Standards
 - Error Prevention
 - Recognition Rather than Recall
 - Flexibility & Efficiency of Use
 - Aesthetic & Minimalist Design
 - Help Users Recognize, Diagnose & Recover from Errors
 - Help & Documentation

Evaluation heuristique : plusieurs experts



1 évaluateur trouve 35% des problèmes d'utilisabilité en 1-2h ; 5 évaluateurs 75%

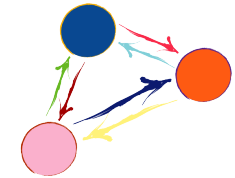
Evaluation heuristique : plusieurs phases



- 1- Mise en contexte préalable : transmettre aux évaluateurs la connaissance sur le domaine et les informations utiles sur les scénarios
- 2- Evaluation individuelle
- 3- Agrégation
- 4- Classification des problèmes par sévérité (priorité)
- 5- Restitution à l'équipe de conception

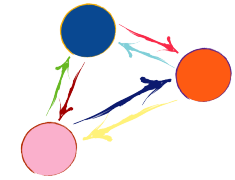
[Bowman, 2002]

Evaluation heuristique : sévérité



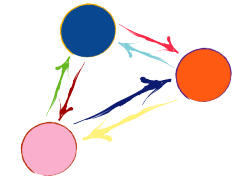
- Echelle des sévérités
 - 0 : ceci n'est pas un problème
 - 1 : problème cosmétique, à régler si on en a le temps
 - 2 : problème mineur, à régler mais pas urgemment
 - 3 : problème majeur, à régler absolument et rapidement
 - 4 : problème catastrophique, sans impasse possible

Promenade cognitive



- Principe : un évaluateur chemine à travers l'interface comme un utilisateur. Ses actions sont basées sur un modèle générique de l'activité
- En conséquence
 - Évalue la capacité de l'IHM à faciliter le processus d'apprentissage exploratoire, i.e. la capacité de l'utilisateur à réaliser des tâches sans formation initiale approfondie
 - Les problèmes liés au domaine échapperont à l'évaluation

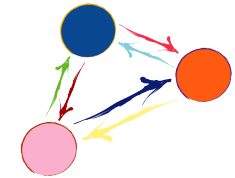
Promenade cognitive : étapes



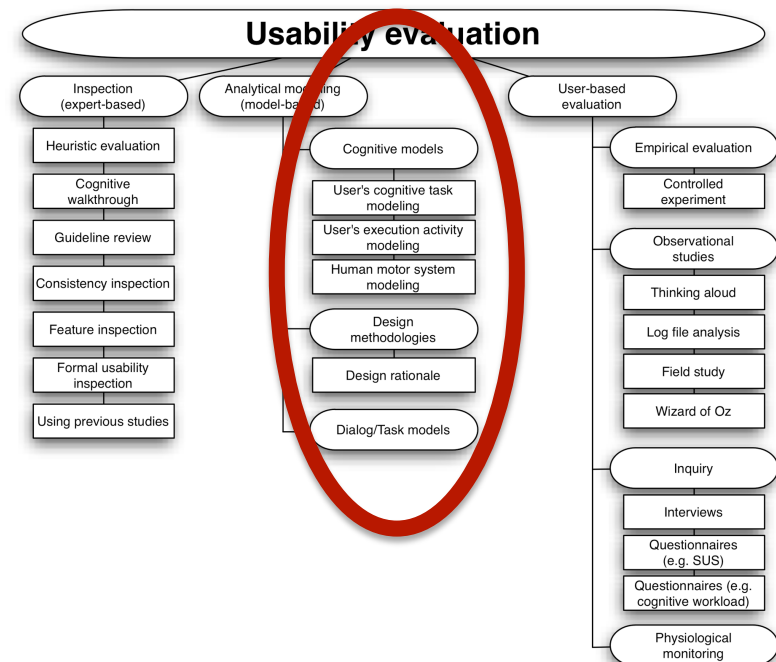
- **Préambule**
 - Description générale des utilisateurs et de leurs connaissances
 - Description spécifique des tâches à réaliser (scénarios)

- **Cheminement**
 - Passer en revue chaque étape des séquences
 - Raconter une histoire crédible expliquant pourquoi l'utilisateur choisirait (ou non) ces actions

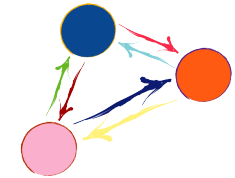
- **Portée : quatre étapes de l'interaction**
 - 1- L'utilisateur se fixe un objectif à atteindre avec le systèmes
 - 2- L'utilisateur recherche dans l'IHM les actions réalisables
 - 3- L'utilisateur choisit l'action la plus appropriée
 - 4- L'utilisateur réalise l'action et évalue le feedback du système



Evaluation par modèles



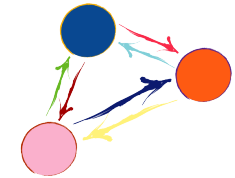
Evaluation par modèles



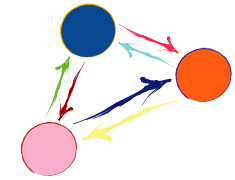
- Principe (Freiberg, 2008)

“As their name suggests, model-based evaluation techniques **use models of interfaces as the basis for the evaluation**. The goal is, to **predict** mostly **quantitative measures** of an interface for example, task duration by **simulating the users' behaviour**. The basic technique consists of 4 steps: describe the interface design in detail, create a model of representative users and their task performance, predict chosen measures by simulating the model, and initially revise or chose the design depending on the prediction. Such a simulation can take place at early stages in the development process and thus valuable usability results can be collected without even implementing a prototype. However, it can be challenging to correctly set up and fine-tune such a model and, even when done, it still might not be a complete or perfected mapping of the actual interface.”

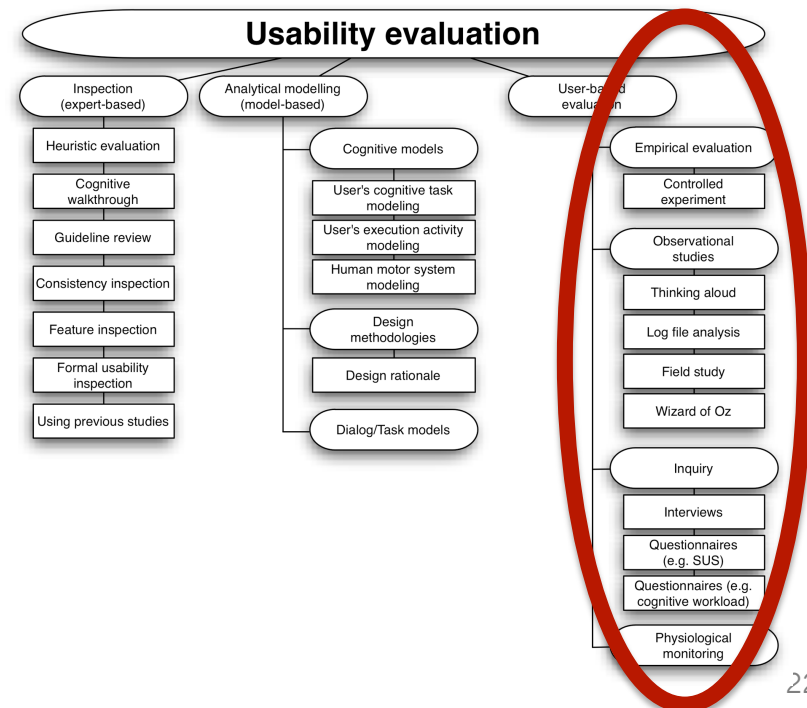
Evaluation par modèles



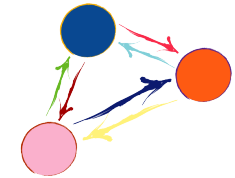
- Trois classes de modèles (Dix, 2004)
 - Cognitive models to predict user's performance (e.g., GOMS (goals, operators, methods and selection), KLM (keystroke-level model))
 - Design methodologies (e.g., design rationale)
 - Dialog models (e.g., state transition networks for unreachable states, circular dialogs and complexity)



Evaluation expérimentale



Expérimentation contrôlée

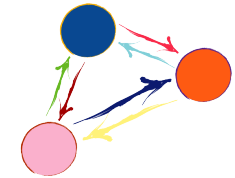


- Principe
 - Expérimentation sur un système fonctionnel
 - Séances contrôlées d'utilisation du système par des sujets (participants)
 - Sujets neutres (i.e., externes à l'équipe de conception) et représentatifs du public ciblé
 - Enregistrement des séances à finalité d'analyse

- Portée
 - Utilité
 - Utilisabilité

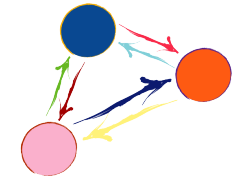
- Etude
 - Qualitative
 - Quantitative

Expérimentation contrôlée



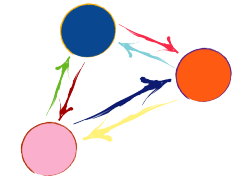
- Organisation
 - 1. Définir ce que l'on veut évaluer
 - 2. Établir un plan de tâches
 - 3. Faire passer le test
 - 4. Analyser les données collectées. Attention à la fiabilité du test (reproductibilité) et à la validité du test (pertinence des tâches, des utilisateurs et des résultats)

Expérimentation contrôlée



- Déroutement du test
 - 1. Préparer le système avant l'arrivée du participant
 - 2. Souhaiter la bienvenue au participant
 - 3. Informer le participant de l'objet du test, et en particulier lui dire que “c'est le système que l'on teste, non pas le participant”. L'assurer de la confidentialité des données. Ne pas l'informer que VOUS êtes concepteur du système, l'encourager à critiquer
 - 4. Lui donner les connaissances minimales pour exécuter la tâche
 - 5. (Lui demander de “penser tout haut”)
 - 6. Durant le test, enregistrer le comportement, les paroles du participant
 - 7. Réaliser un entretien post-test pour recueillir les commentaires généraux du participant et lui poser les questions que vous avez préparées.

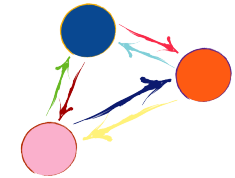
Expérimentation contrôlée qualitative



- Exploration libre
 - Donner le contexte au participant (raison d'être du système)
 - Le laisser explorer librement le système
 - Lui demander de reproduire une activité récente vécue, en lien avec le système

- Plan de tâches (tâches principales – cf arbre des tâches)
 - Choisir des tâches centrées utilisateur (ex: “faites vos courses”), non pas des tâches systèmes (ex: “créez un compte”)
 - Exemple
 - Trouver le prix de l'article X
 - Quel est le meilleur rapport poids/prix pour un article de type Y
 - Commander les articles Z, W et K
 - Trouvez la réponse à la question Q

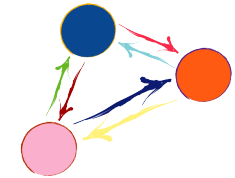
Expérimentation contrôlée qualitative



- Modération

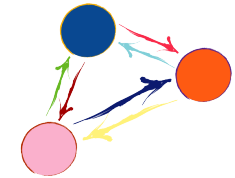
	Penser tout haut	Questions de l'expérimentateur
Pendant la session	<ul style="list-style-type: none">+ Détection des problèmes en temps réel+ Observations des stratégies de correction- Interférence avec la tâche (performances)- Difficulté à entretenir	<ul style="list-style-type: none">+ Facile à entretenir+ Pas de problème de mémorisation- Interférence avec la tâche- Interruption des actions du participant- Certains problèmes peuvent passer inaperçus
Après la session	<ul style="list-style-type: none">+ Pas d'interférence avec la tâche- Mémorisation difficile, surtout pour de longues sessions	<ul style="list-style-type: none">+ Pas d'interférence avec la tâche- Risque d'oubli des événements mentionnés- Certains problèmes peuvent passer inaperçus

Expérimentation contrôlée qualitative

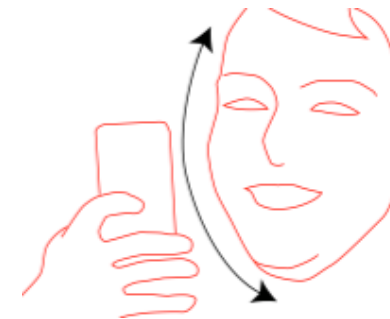


- Exploitation des résultats : si problème
 - Systématiquement le noter
 - Évaluer la fiabilité du problème
 - Identifier sa cause
 - Imaginer une modification de conception
 - Évaluer le risque de la modification (i.e. introduction d'autres problèmes)

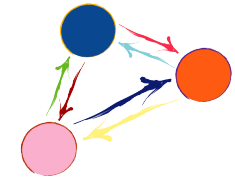
Expérimentation contrôlée quantitative



- Choix entre deux variantes d'IHM
 - Évaluation statistique des effets de variations de l'IHM sur l'usage du système.
 - Exemple : “Le contrôle du défilement par rotation de la tête est-il supérieur au contrôle avec un doigt ?”

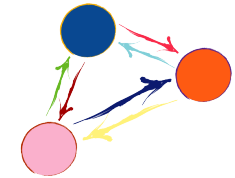


Expérimentation contrôlée quantitative



- Plan d'expérimentation
 - Choix d'une tâche mettant en oeuvre l'interaction à tester
 - Exemple : Trouver le numéro de téléphone de 10 contacts
 - Identification des variables indépendantes, i.e. contrôlées (Interaction de défilement)
 - Identification des conditions (rotation du visage (A) / glissé du doigt (B))
 - Choix des variables dépendantes, i.e. mesurées (temps d'accomplissement de la tâche, taux d'erreur)
 - Évaluation statistique des effets de variations de l'IHM sur l'usage du système

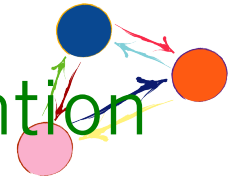
Expérimentation contrôlée quantitative



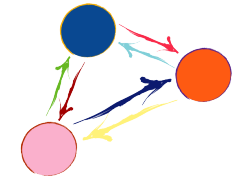
- Variabilité
 - Maximiser les variations entre les participants (à l'intérieur de la catégorie cible) → tester un échantillon représentatif
 - Minimiser les variations entre les conditions → mesurer uniquement l'effet de la variable indépendante → Fiche de consigne

- Analyse des données : test d'hypothèses
 - Répondre à la question : quelle est la probabilité que l'écart mesuré soit la cause du hasard ?
 - La différence est significative ?

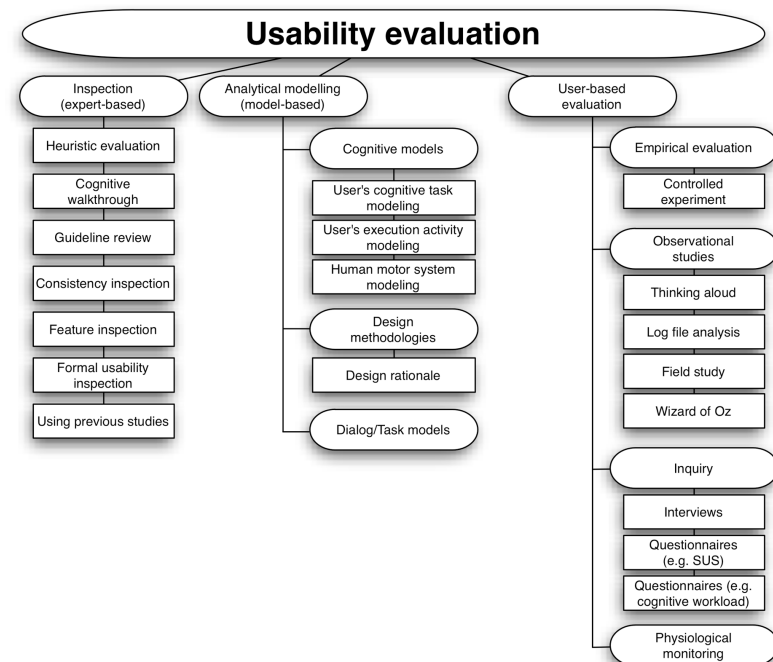
Expérimentation contrôlée : points d'attention



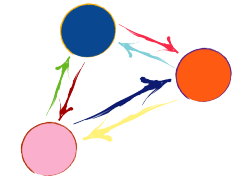
- Recrutement des participants : les plus représentatifs possibles ...
- Formulation d'une hypothèse : en termes de variables
- Deux types de variables
 - Indépendantes : contrôlées
 - Dépendantes : mesurées
- Hypothèse : si évolution de v_ind_i , prédiction de v_dep_j



Complémentarité

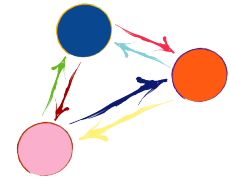


Critères de choix



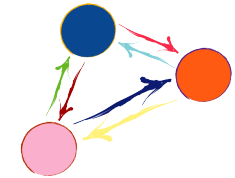
- Huit critères (Dix, 2004)
 - Design vs. implementation
 - Laboratory vs. field study: laboratory studies allow controlled experimentation and observation while losing something of the naturalness of the user's environment. Field studies retain the latter but do not allow control over user activity.
 - Subjective vs. objective: Evaluation techniques also vary according to their objectivity – some techniques rely heavily on the interpretation of the evaluator, others would provide similar information for anyone correctly carrying out the procedure. The more subjective techniques, such as cognitive walkthrough or think aloud, rely to a large extent on the knowledge and expertise of the evaluator, who must recognize problems and understand what the user is doing. Ideally, both objective and subjective approaches should be used.

Critères de choix



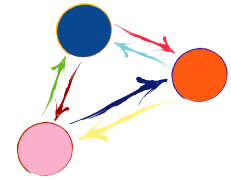
- Huit critères (Dix, 2004)
 - Design vs. implementation
 - Laboratory vs. field study
 - Subjective vs. objective
 - Qualitative vs. quantitative measures: the former is usually numeric and can be easily analyzed using statistical techniques. The latter is non-numeric and is therefore more difficult to analyze, but can provide important detail that cannot be determined from numbers. The type of measure is related to the subjectivity or objectivity of the technique, with subjective techniques tending to provide qualitative measures and objective techniques, quantitative measures.
 - Information provided: the information required by an evaluator at any stage of the design process may range from low-level information to enable a design decision to be made (for example, which font is most readable) to higher-level information, such as ‘Is the system usable?’

Critères de choix



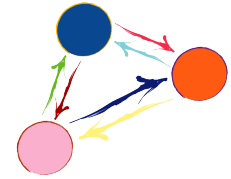
- Huit critères (Dix, 2004)
 - Design vs. implementation
 - Laboratory vs. field study
 - Subjective vs. objective
 - Qualitative vs. quantitative measures
 - Information provided
 - Immediacy of the response
 - Intrusiveness: certain techniques, particularly those that produce immediate measurements, are obvious to the user during the interaction and therefore run the risk of influencing the way the user behaves. Sensitive activity on the part of the evaluator can help to reduce this but cannot remove it altogether. Most immediate evaluation techniques are intrusive, with the exception of automatic system logging. Unfortunately, this is limited in the information that it can provide and is difficult to interpret

Critères de choix



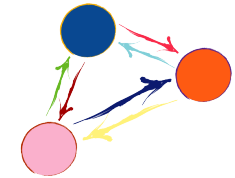
- Huit critères (Dix, 2004)
 - Design vs. implementation
 - Laboratory vs. field study
 - Subjective vs. objective
 - Qualitative vs. quantitative measures
 - Information provided
 - Immediacy of the response
 - Intrusiveness
 - Resources required: the final consideration when selecting an evaluation technique is the availability of resources. Resources to consider include equipment, time, money, participants, expertise of evaluator and context. Some decisions are forced by resource limitations: it is not possible to produce a video protocol without access to a video camera (and probably editing facilities as well). However, other decisions are not so clear cut.”

Critères de choix



- Huit critères (Dix, 2004)
 - Design vs. implementation
 - Laboratory vs. field study
 - Subjective vs. objective
 - Qualitative vs. quantitative measures
 - Information provided
 - Immediacy of the response
 - Intrusiveness
 - Resources required

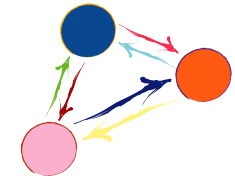
Critères de choix



- Caractérisation des techniques analytiques (Dix, 2004)

	Cognitive walkthrough	Heuristic evaluation	Review based	Model based
Stage	Throughout	Throughout	Design	Design
Style	Laboratory	Laboratory	Laboratory	Laboratory
Objective?	No	No	As source	No
Measure	Qualitative	Qualitative	As source	Qualitative
Information	Low level	High level	As source	Low level
Immediacy	N/A	N/A	As source	N/A
Intrusive?	No	No	No	No
Time	Medium	Low	Low-medium	Medium
Equipment	Low	Low	Low	Low
Expertise	High	Medium	Low	High

Critères de choix



- Caractérisation des techniques (Dix, 2004)

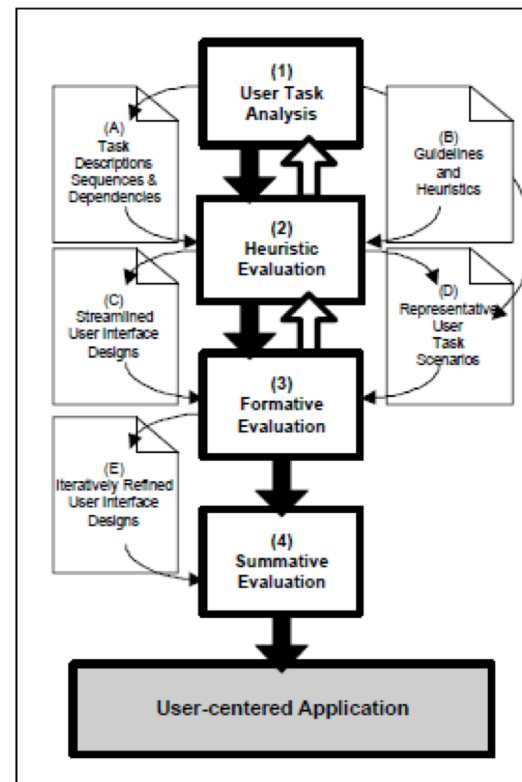
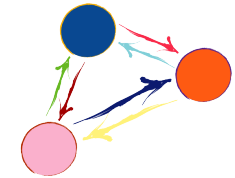
	Cognitive walkthrough	Heuristic evaluation	Review based	Model based
Stage	Throughout	Throughout	Design	Design
Style	Laboratory	Laboratory	Laboratory	Laboratory
Objective?	No	No	As source	No
Measure	Qualitative	Qualitative	As source	Qualitative
Information	Low level	High level	As source	Low level
Immediacy	N/A	N/A	As source	N/A
Intrusive?	No	No	No	No
Time	Medium	Low	Low-medium	Medium
Equipment	Low	Low	Low	Low
Expertise	High	Medium	Low	High

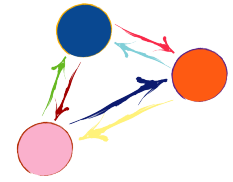
	Experiment	Interviews	Questionnaire
Stage	Throughout	Throughout	Throughout
Style	Laboratory	Lab/field	Lab/field
Objective?	Yes	No	No
Measure	Quantitative	Qualitative/ quantitative	Qualitative/ quantitative
Information	Low/high level	High level	High level
Immediacy	Yes	No	No
Intrusive?	Yes	No	No
Time	High	Low	Low
Equipment	Medium	Low	Low
Expertise	Medium	Low	Low

	Think aloud ¹	Protocol analysis ²	Post-task walkthrough
Stage	Implementation	Implementation	Implementation
Style	Lab/field	Lab/field	Lab/field
Objective?	No	No	No
Measure	Qualitative	Qualitative	Qualitative
Information	High/low level	High/low level	High/low level
Immediacy	Yes	Yes	No
Intrusive?	Yes	Yes ³	No
Time	High	High	Medium
Equipment	Low	High	Low
Expertise	Medium	High	Medium

	Eye tracking	Physiological measurement
Stage	Implementation	Implementation
Style	Lab	Lab
Objective?	Yes	Yes
Measure	Quantitative	Quantitative
Information	Low level	Low level
Immediacy	Yes	Yes
Intrusive?	No ¹	Yes
Time	Medium/high	Medium/high
Equipment	High	High
Expertise	High	High

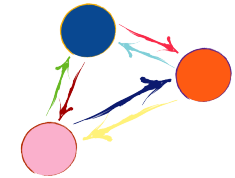
Evaluation séquentielle



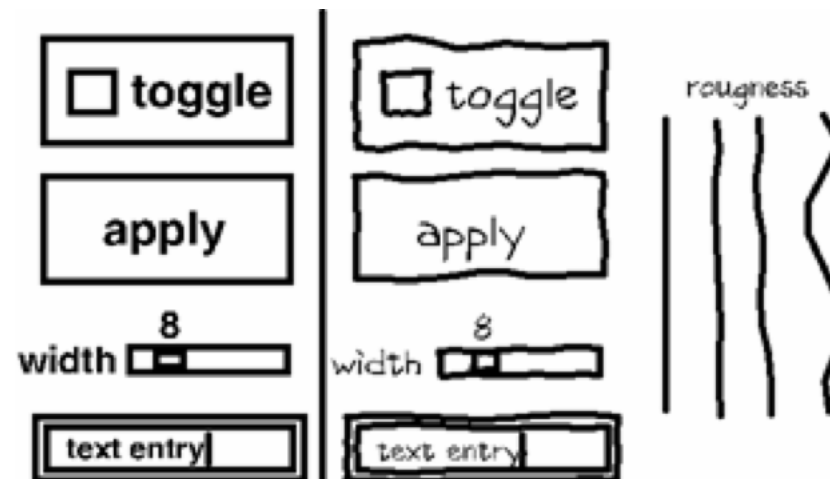


En pratique

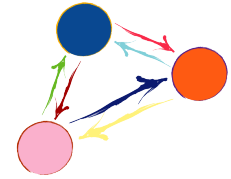
Principes gagnants



- #1. Evaluation heuristique avant expérimentale : évacuer les problèmes d'utilisabilité au plus tôt !
- #2. Evaluation en basse fidélité : convaincre qu'il est encore temps d'améliorer ! [Meyer 1996] [Meyer 2005]



Principes gagnants



- #1. Evaluation heuristique avant expérimentale : évacuer les problèmes d'utilisabilité au plus tôt !
- #2. Evaluation en basse fidélité : convaincre qu'il est encore temps d'améliorer ! [Meyer 1996] [Meyer 2005]
- #3. Evaluation comparative : tester plusieurs IHM est mieux qu'une seule ! [Tohidi 2006]

Principes gagnants

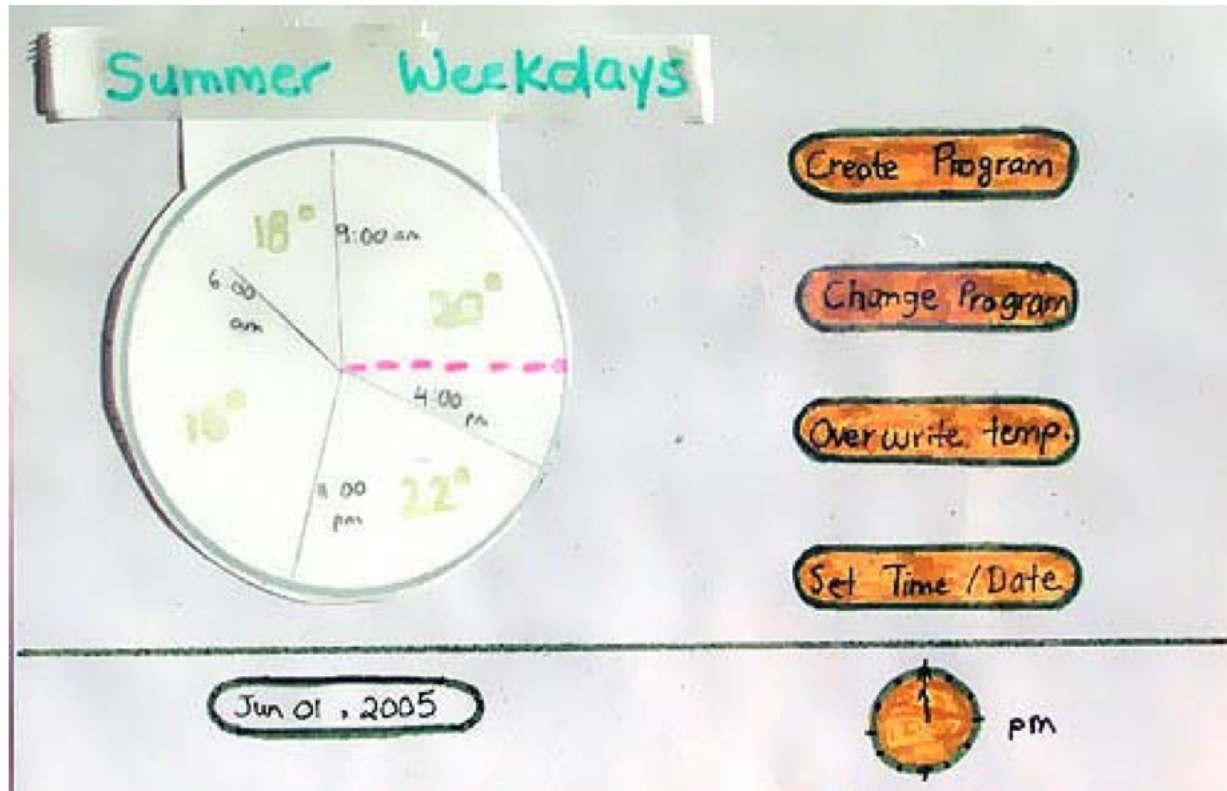
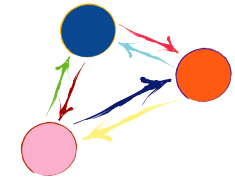


Figure 1. The “Circular” paper prototype

Principes gagnants

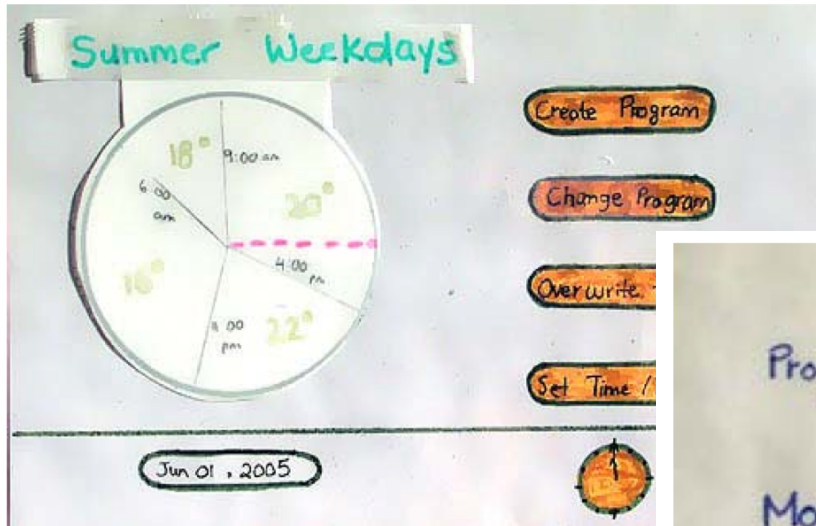
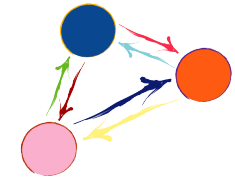


Figure 1. The “Circular” paper prototype

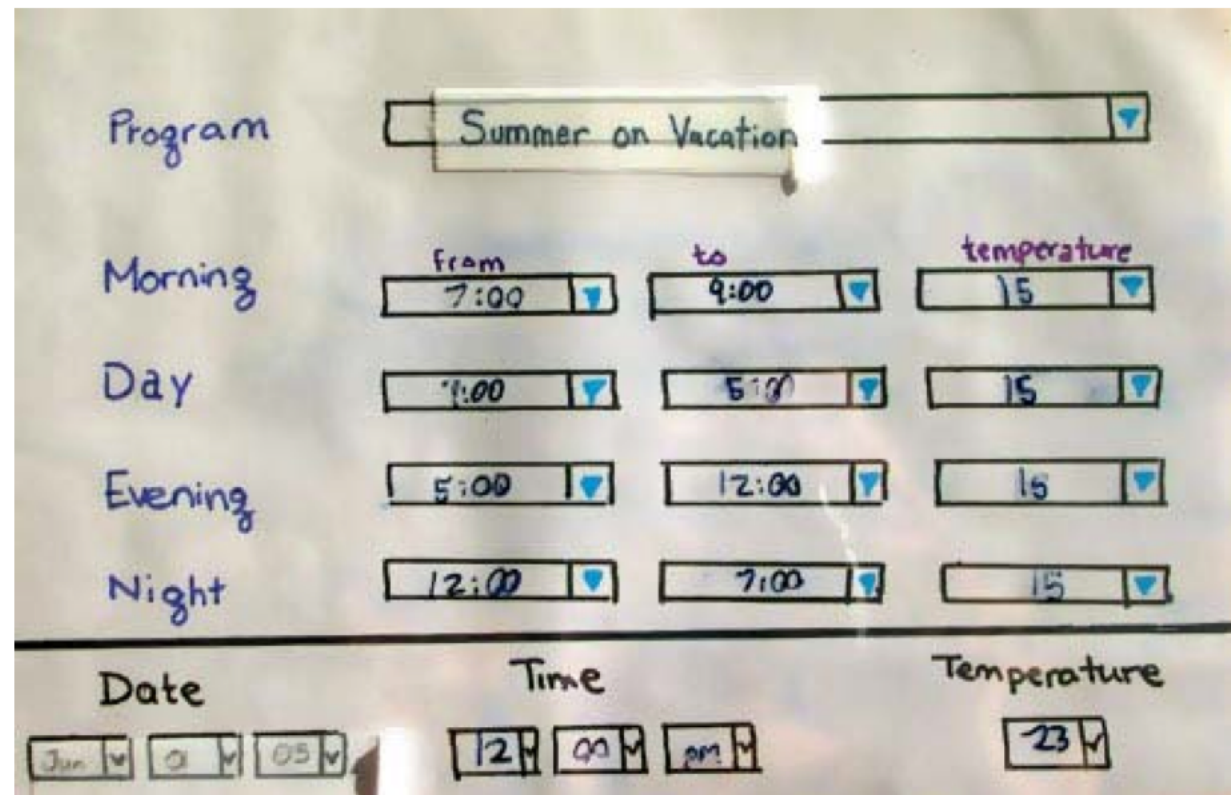


Figure 2. The “Tabular” paper prototype

Principes gagnants

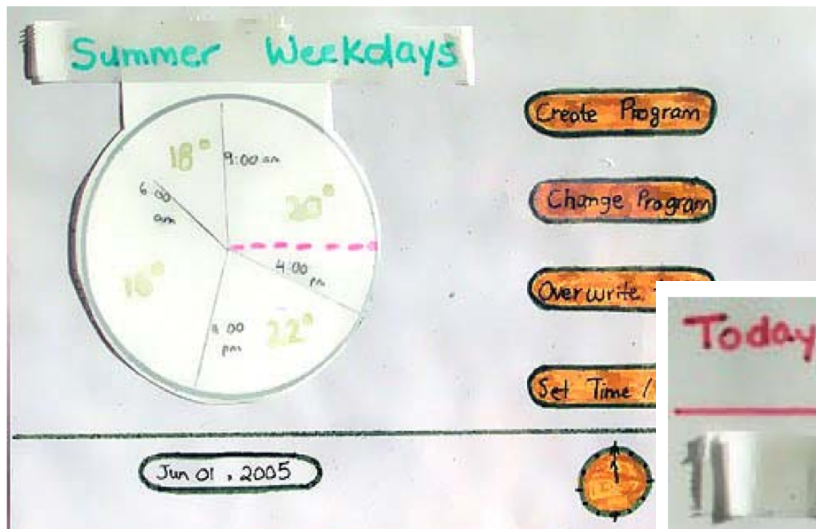
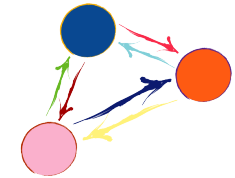


Figure 1. The “Circular” paper prototype

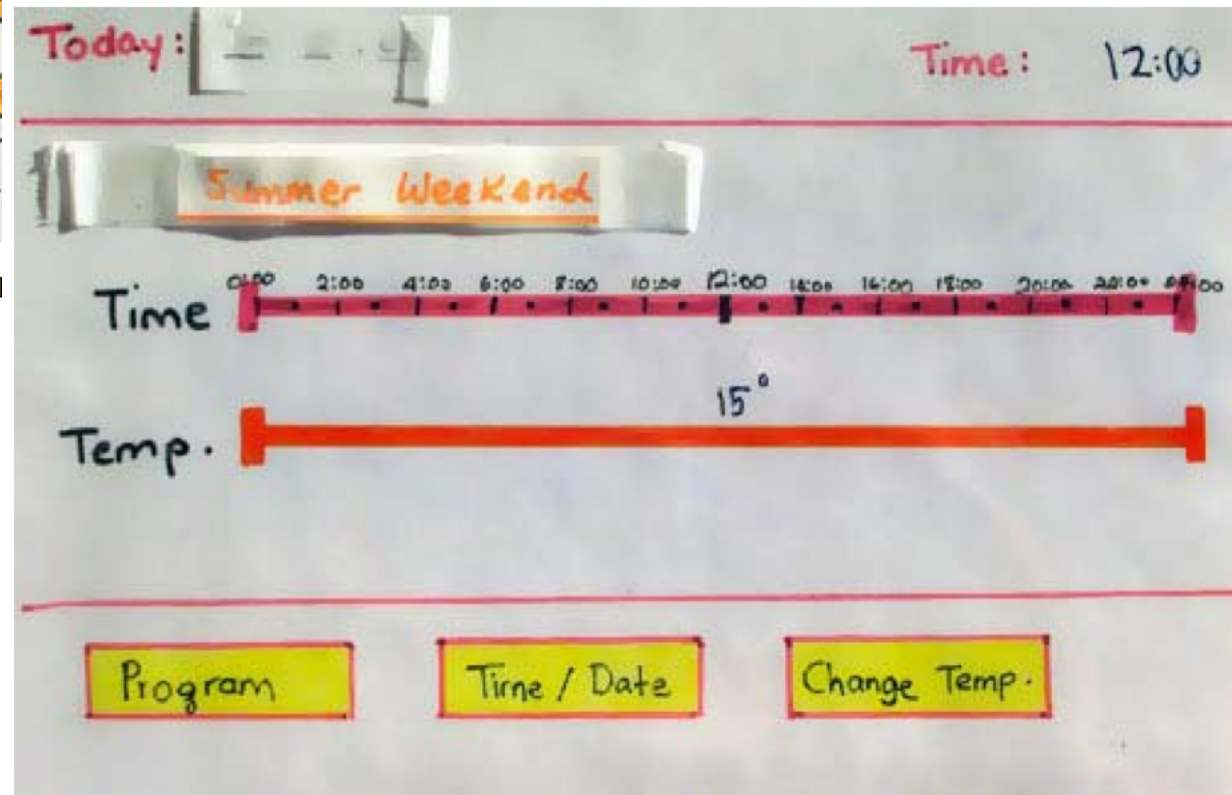
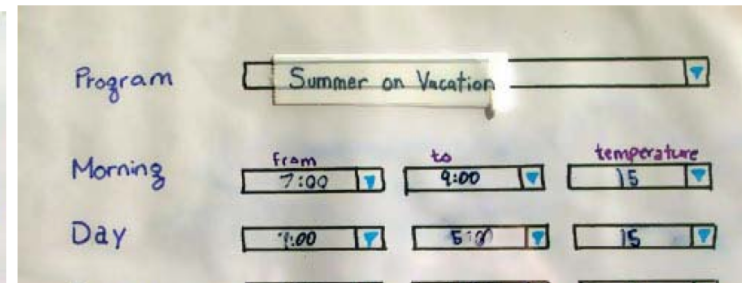


Figure 3. The “Linear” paper prototype
[Tohidi 2006]

Principes gagnants

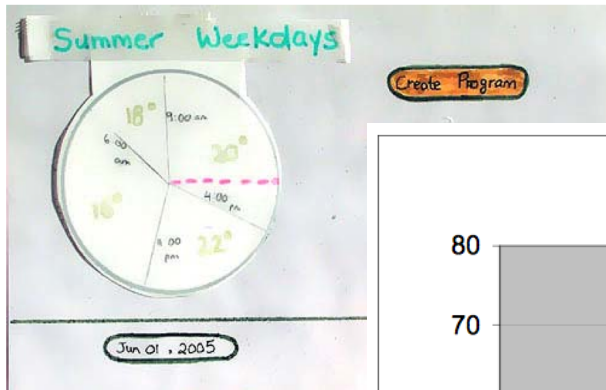
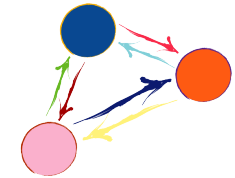
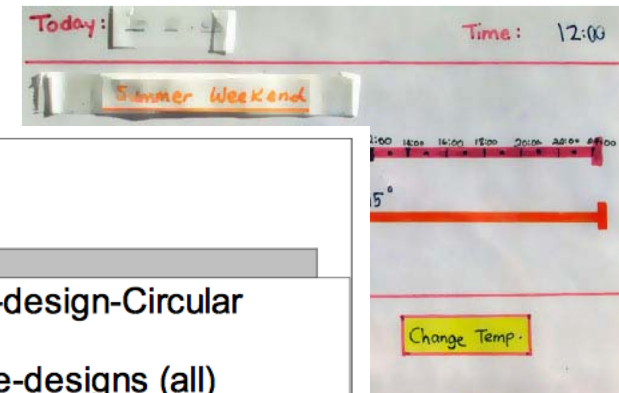
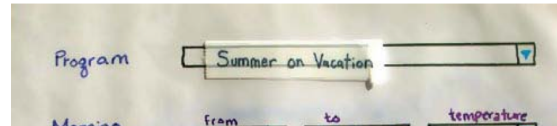


Figure 1. The "Circular" prototype



paper prototype

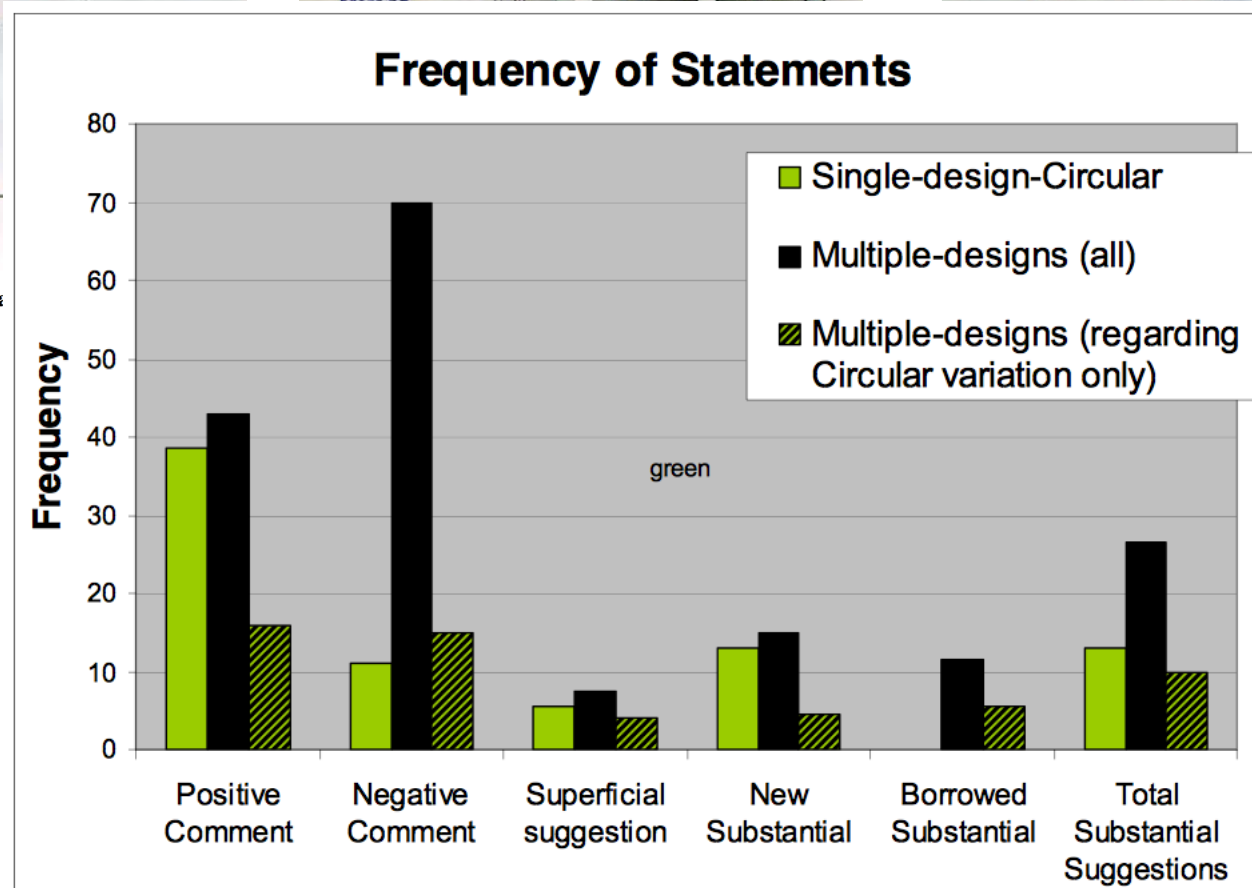


Figure 5. Frequency of statements for Circular prototype

Principes gagnants

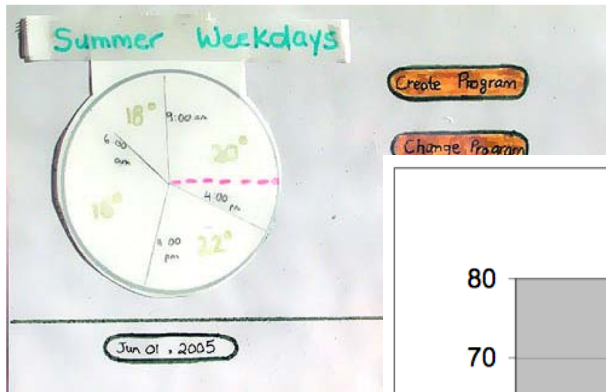
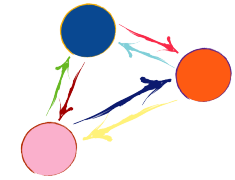
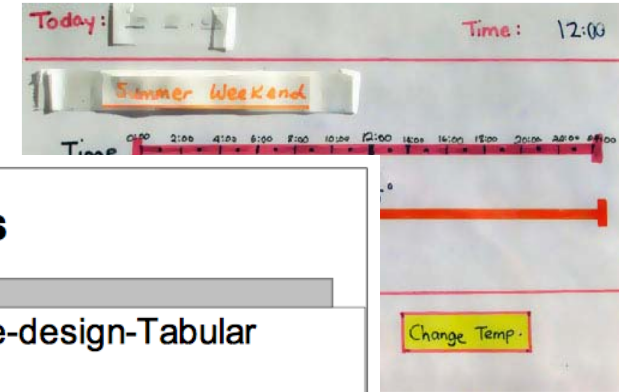
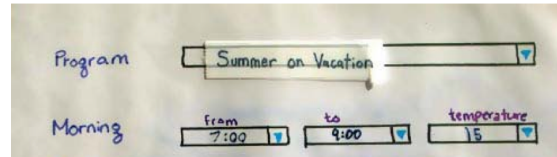


Figure 1. The “Circular” pap



aper prototype

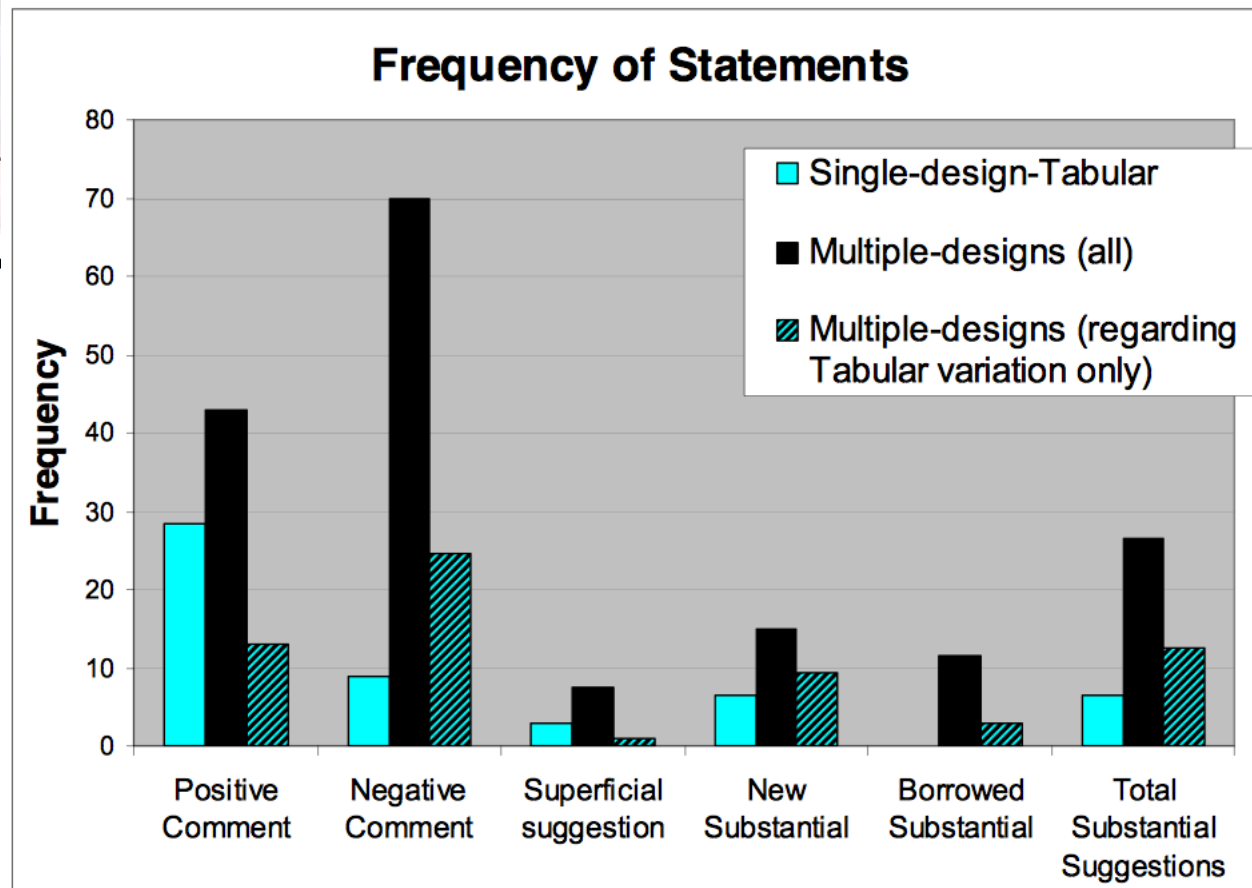


Figure 6. Frequency of statements for Tabular prototype

Principes gagnants

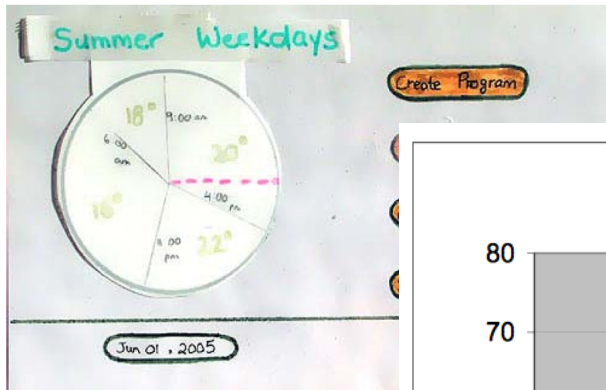
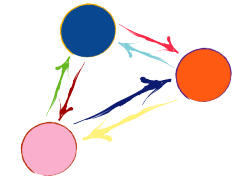
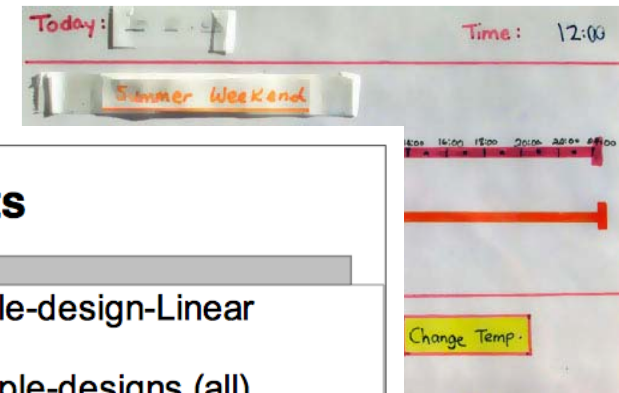
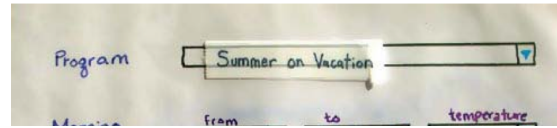


Figure 1. The "Circular" paper



er prototype

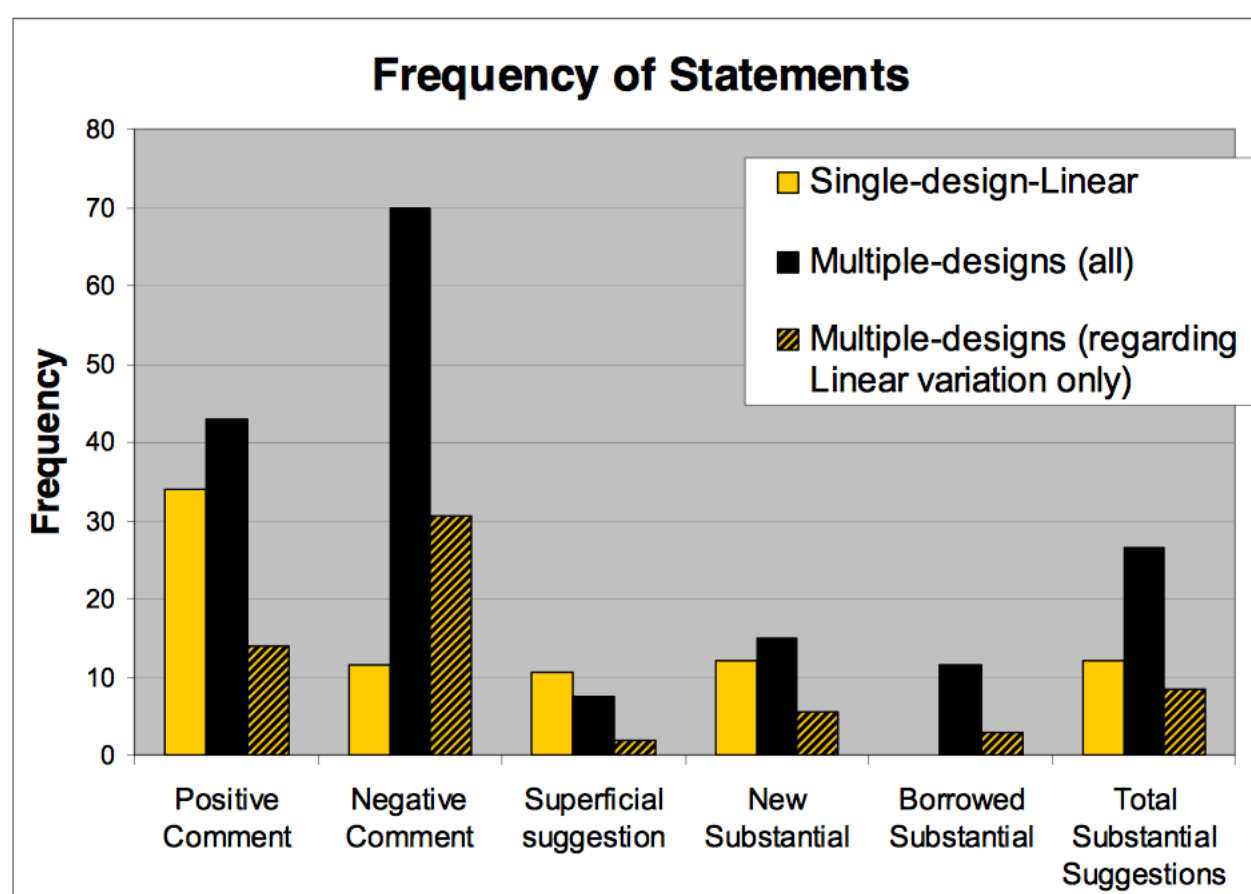


Figure 7. Frequency of statements for Linear prototype

[Tohidi 2006]

Principes gagnants

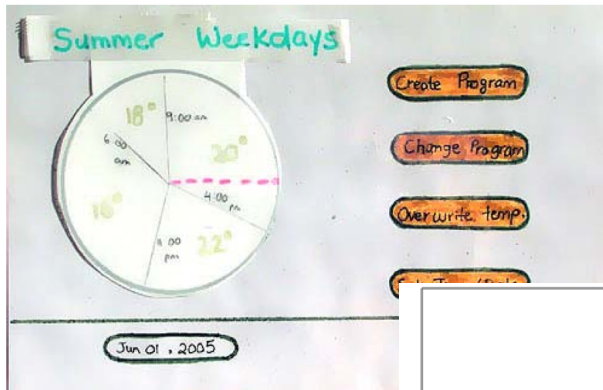
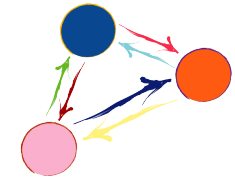
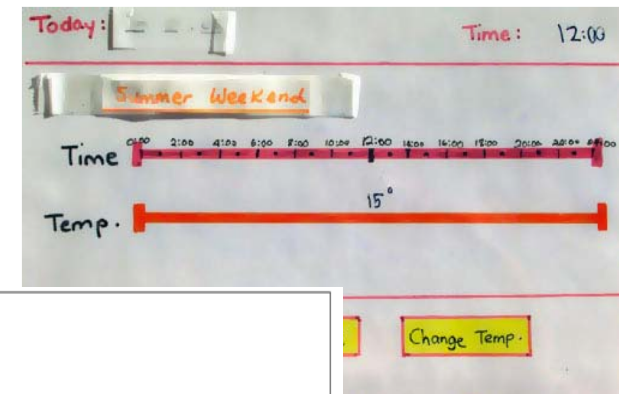
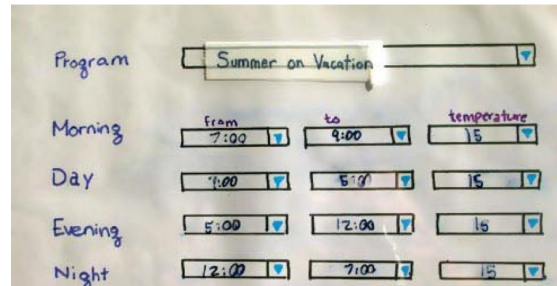


Figure 1. The “Circular” paper



” paper prototype

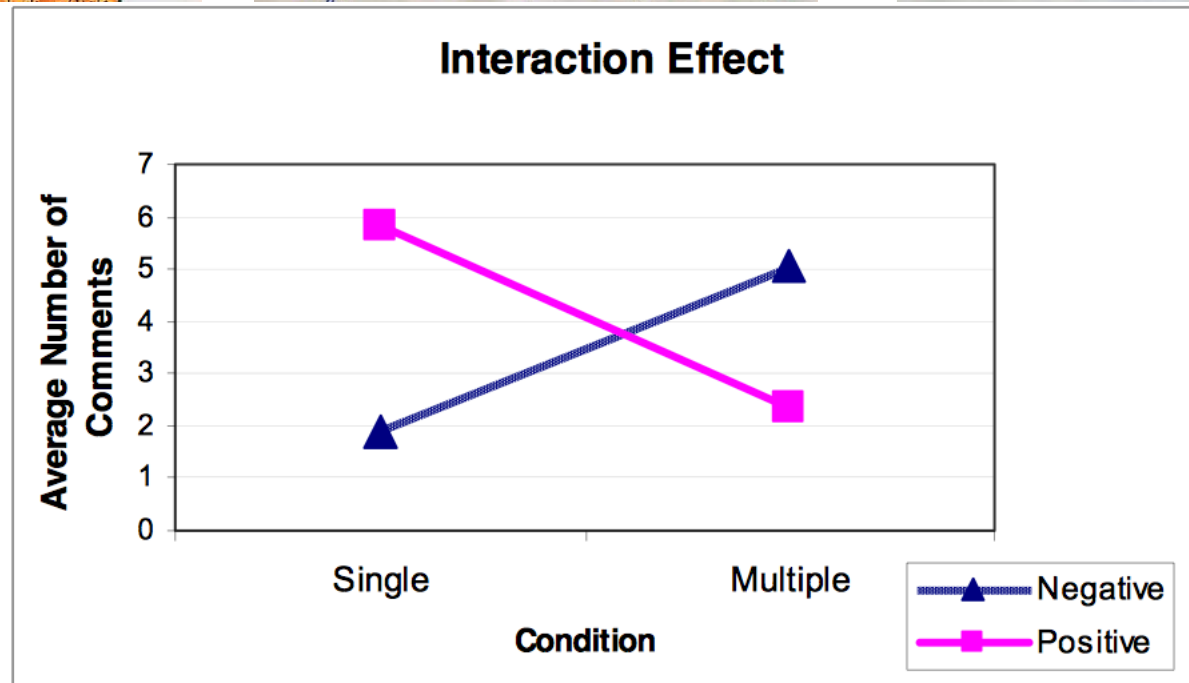
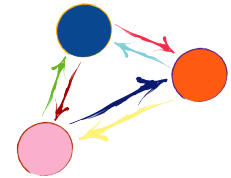
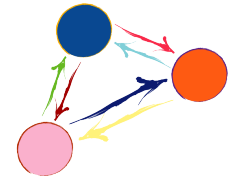


Figure 8. Interaction between positive and negative comments made in response to the Linear design in both conditions.

Principes gagnants

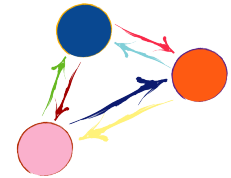


- #1. Evaluation heuristique avant expérimentale : évacuer les problèmes d'utilisabilité au plus tôt !
- #2. Evaluation en basse fidélité : convaincre qu'il est encore temps d'améliorer ! [Meyer 1996] [Meyer 2005]
- #3. Evaluation comparative : tester plusieurs IHM est mieux qu'une seule ! [Tohidi 2006]
- #4. Etre TRES précautionneux sur ce qu'on « mesure » (ex : learning effect)
- #5. Etre TRES précautionneux sur les éventuels biais (ex : S1 puis S2 pour 50% des sujets ; S2 puis S1 pour les autres 50%)
- #6. Tendance aux évaluations in the wild, longitudinales



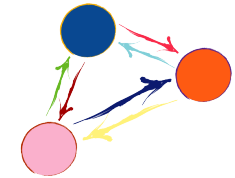
Un exemple

Cas d'étude [Harries 2013]

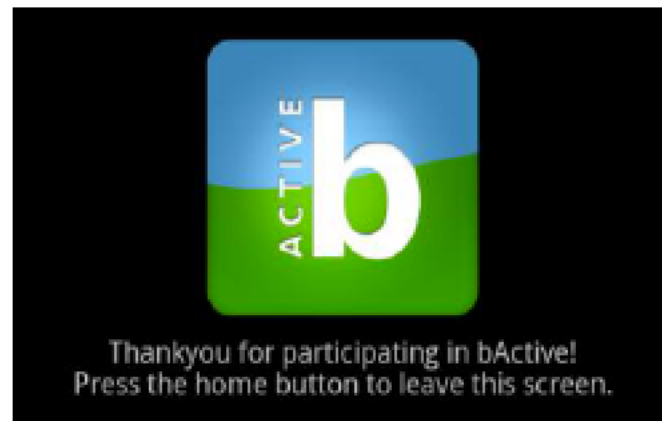


- Domaine : technologies persuasives pour la santé
- Sujet : marche
- Titre : « Walking in the Wild – Using an Always-on Smartphone Application to Increase Physical Activity »

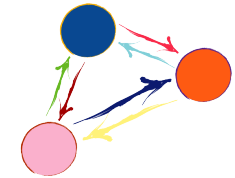
Cas d'étude [Harries 2013]



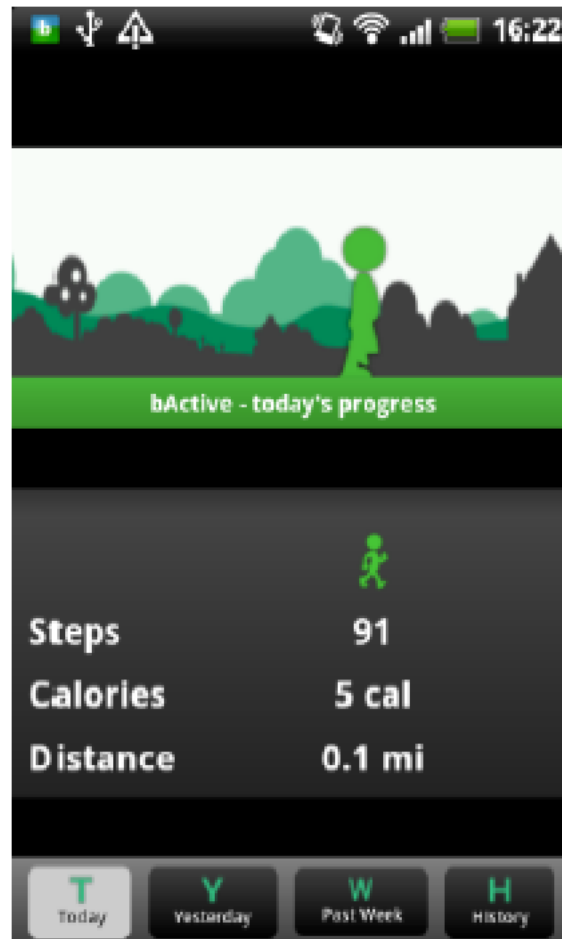
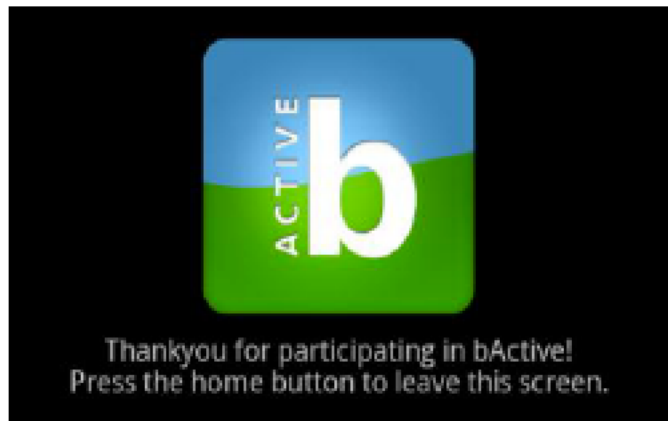
- Domaine : technologies persuasives pour la santé
- Sujet : marche
- Titre : « Walking in the Wild – Using an Always-on Smartphone Application to Increase Physical Activity »



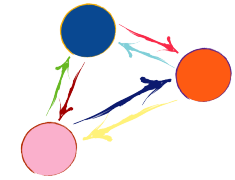
Cas d'étude [Harries 2013]



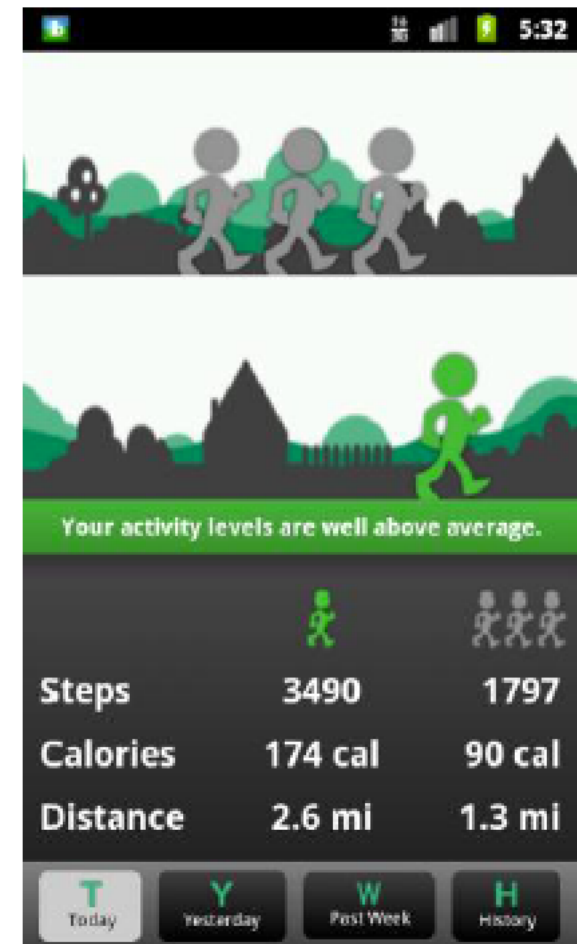
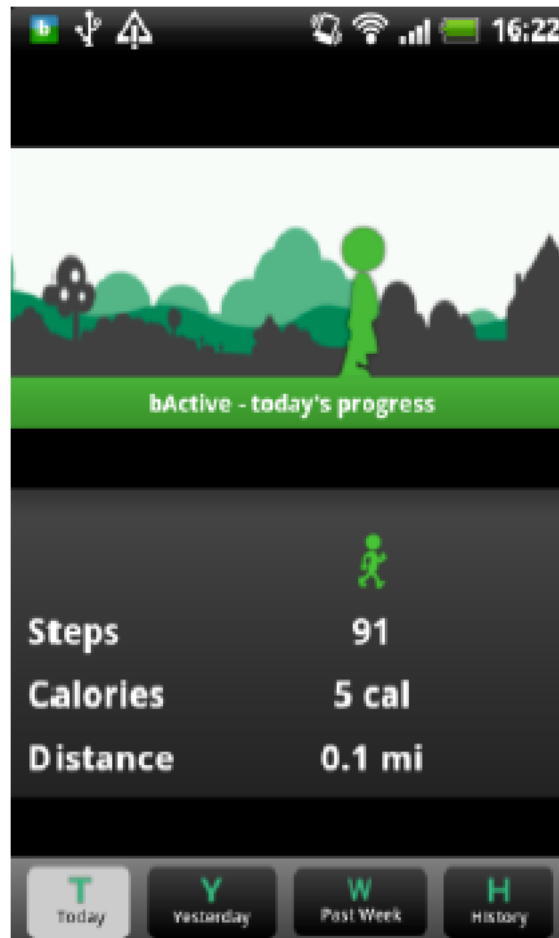
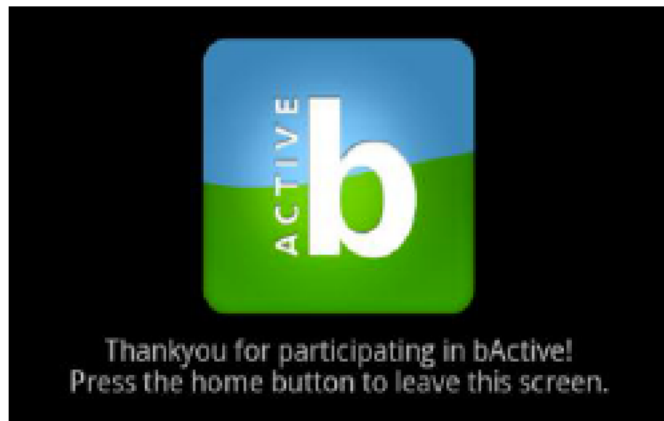
- Domaine : technologies persuasives pour la santé
- Sujet : marche
- Titre : « Walking in the Application to Increase



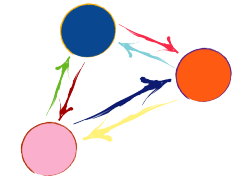
Cas d'étude [Harries 2013]



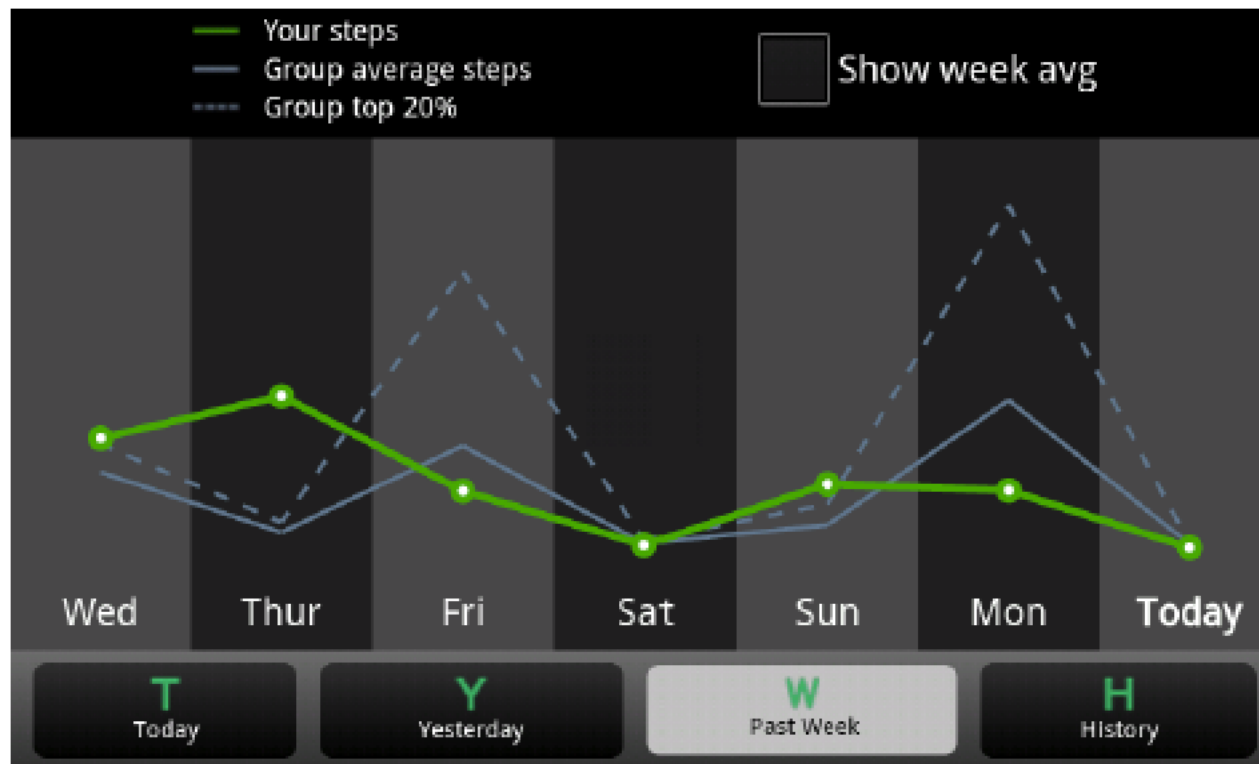
- Domaine : technologies persuasives pour la santé
- Sujet : marche
- Titre : « Walking in the City » Application to Increase Physical Activity



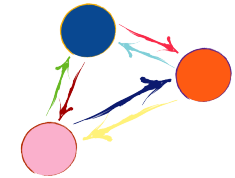
Cas d'étude [Harries 2013]



- Domaine : technologies persuasives pour la santé
- Sujet : marche
- Titre : « Walking in the Wild – Using an Always-on Smartphone Application to Increase Physical Activity »

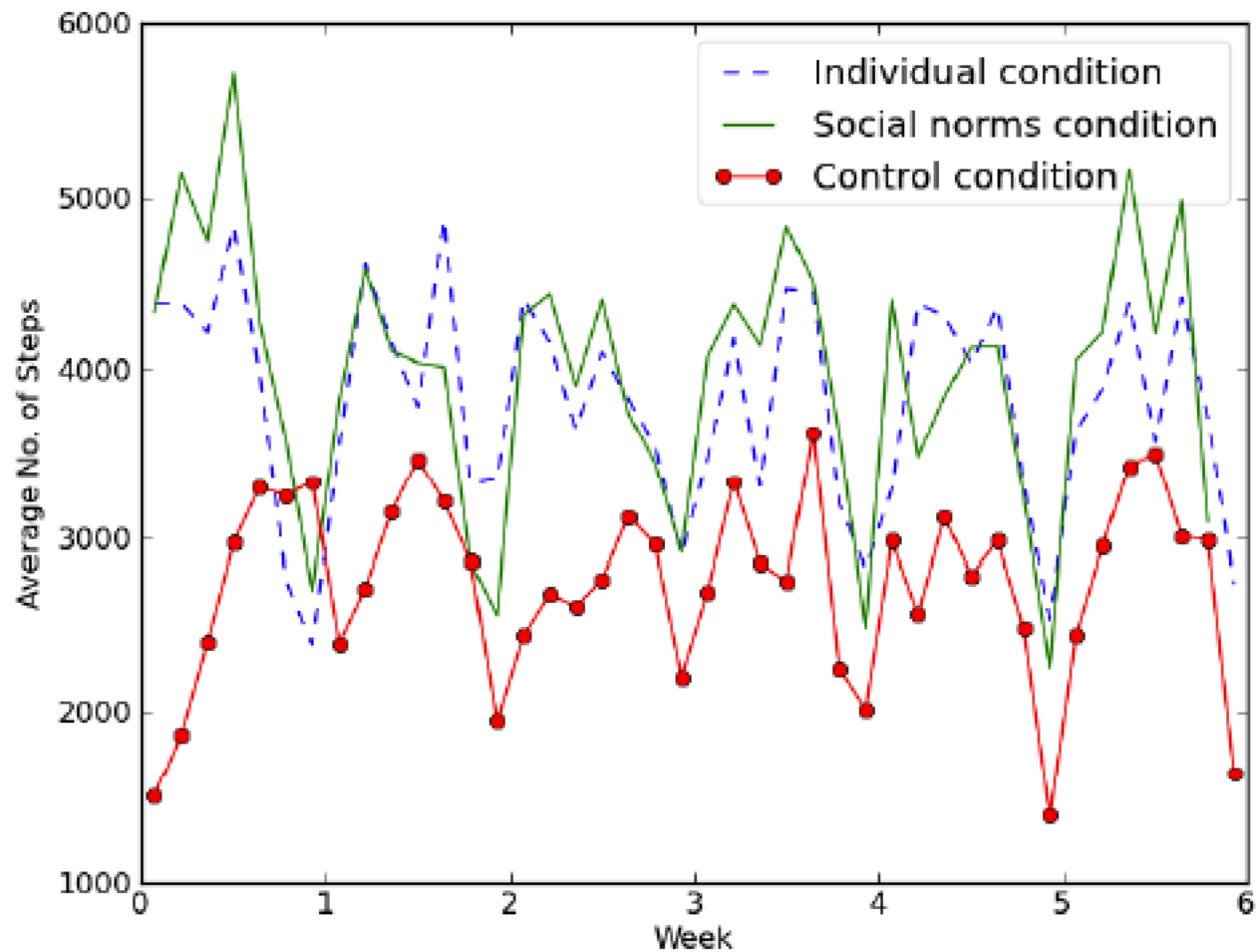


Cas d'étude [Harries 2013]

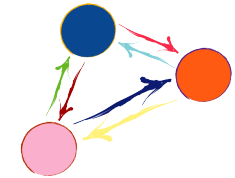


- Domaine : technologies persuasives pour la santé

▪ S
▪ Ti
▪ A



Cas d'étude [Harries 2013]



- Domaine : technologies persuasives pour la santé
- Sujet : marche
- Titre : « Walking in the Wild – Using an Always-on Smartphone Application to Increase Physical Activity »

