



UNIVERSITÀ
DEGLI STUDI
DI BRESCIA

Interazione Persona-Calcolatore

Metodi di raccolta dati

Prof.ssa Daniela Fogli

Dipartimento di Ingegneria dell'Informazione

Raccolta dati (oltre l'esperimento)

- **Osservazione**

- Diretta (di persona)
- Indiretta (con registrazioni)
- Protocolli verbali

- Raccolta delle **opinioni** dell'utenza

- Interviste
- Questionari
- Focus group

- **Monitoraggio**

- Logging dell'uso del sistema

Utile combinare le varie tecniche

L'osservazione nell'ingegneria dell'usabilità

- L'osservazione dell'utente al lavoro è un metodo importante di raccolta dati, sia per **l'analisi dei compiti** sia per **la valutazione di usabilità** dei sistemi installati
- Usati negli studi sul campo (**field studies** o **in-the-wild studies**)
- L'osservazione può essere condotta con varie tecniche, cercando di essere **meno intrusivi possibile**, compatibilmente con il metodo scelto
- Gli **scopi dettagliati** (es. scoprire gli errori nell'eseguire un'attività, etc.) e le metriche dell'osservazione non sono necessariamente definiti a priori
- Si possono scoprire **modi d'uso inaspettati**
- Occorre far firmare un **consenso informato**

Osservazione degli utenti: considerazioni

- Lo scopo di questa valutazione è permettere ai progettisti di capire come gli utenti usano il sistema **nel loro ambiente naturale** e come il sistema **si integra** con le altre attività (si parla anche di **metodi etnografici**) – sul lavoro, ma anche nella vita familiare (e.g., smart home)
- È utile **combinare** più modi di osservare (**triangulation**)
- È comune una qualche forma di **partecipazione** degli utenti nella raccolta, analisi o interpretazione dei dati
 - Protocolli verbali durante l'osservazione diretta
 - Diari degli utenti
 - Analisi delle registrazioni a seguito di osservazione indiretta

Protocolli verbali

- Fuoco dell'attenzione su **ciò che l'utente dice**
- Due modalità:
 - Raccogliere **i commenti che l'utente fa spontaneamente** (in parallelo al video o all'osservazione dell'osservatore)
 - Si capiscono: intenzioni, piani, il linguaggio dell'utente (i nomi delle attività, degli oggetti, degli strumenti)
 - Poco invasivo
 - Poco adatto con utenti silenziosi
 - "**Think-aloud**" o "**Thinking aloud**" (pensare ad alta voce): si chiede all'utente di commentare ciò che sta facendo
 - Raccolta di dati qualitativi importanti
 - Carico cognitivo sull'utente
 - Si scoprono le teorie 'ingenue' dell'utente sulle cause e i rimedi

Come si conduce il “Think-aloud”

- Lo sperimentatore stimola l'utente a parlare
- Lo sperimentatore non risponde alle richieste dell'utente o risponde secondo norme prefissate
- Per ridurre il disagio lo sperimentatore fa prima una **demo** all'utente
- Oppure si usano **copie di utenti**, tipicamente un esperto che insegna a un novizio a usare il sistema

Pro e contro dei protocolli verbali

- **Effetto Hawthorne**: modo di agire non naturale
- Utile per il progettista ma **difficile interpretare i dati**
- Utile per l'utente (dovendo spiegare a volte individua errori nei suoi modelli mentali) ma **carico cognitivo**
- Un utile complemento (o alternativa) è il **“post-event protocol”**: l'utente commenta a posteriori ciò che ha fatto e ciò che ha detto

Alcuni risultati inaspettati

- Pensare ad alta voce permette all'utente
 - di trovare a volte inconsistenze
 - nella sua attività
 - nel modello del compito
 - nell'individuazione di parti critiche di compito/sistema
 - di migliorare le sue prestazioni
 - di prevenire errori

Raccogliere le opinioni degli utenti

- **Intervista:** dà indicazioni qualitative
- **Questionario:** più verso il quantitativo – con attenzione! – ma meno flessibile
- Sono strumenti di **semantizzazione progressiva**
- **Problemi comuni:** sono opinioni soggettive dell'utente, individuano situazioni di disagio o insoddisfazione
- Danno l'idea di **come l'utente pensa che stiano le cose** NON di come stanno veramente

Interviste e questionari: caratteristiche generali

- Metodi simili perché coinvolgono gli utenti facendo loro **domande**
- I **questionari** sono **su carta o interattivi** su computer, possono essere gestiti senza bisogno di altre persone che facciano le domande agli utenti
- Le **interviste** invece prevedono la **presenza di un intervistatore** che legge le domande all'intervistato e le risposte sono solitamente registrate

Interviste e questionari: caratteristiche generali (2)

- Le interviste generano dei **risultati immediati**
- I risultati dei questionari subiscono **ritardi** dovuti a invio/ricevimento domande/risposte
- Coi questionari, occorrono **meccanismi per invogliare l'utente a rispondere**, ma il vantaggio è che in linea di principio è possibile **inviare un questionario all'intera popolazione interessata**

Un problema

- Sia con interviste che con questionari il problema è **interpretare i risultati**
- **Una causa:** la gente risponde ciò che pensa di dover rispondere (per non fare brutta figura, per non apparire insufficiente, per non offendere l'intervistatore)
- **Esempio:** alla richiesta di quanto tempo l'utente ha speso per cercare la soluzione nella documentazione prima di chiedere al servizio assistenza tutti gli utenti tendono a **sovrestimare** il tempo speso
- Quindi non sempre bisogna interpretare i risultati alla lettera

Questionario

- Delle **interviste** abbiamo già parlato a proposito dell'analisi dei compiti, ma possono essere progettate anche per **valutare usabilità**
- **Questionario:** un insieme ordinato di domande **non ambigue**
- **Precio:** può raggiungere un gran numero di utenti (anche quelli impossibilitati a partecipare a un'intervista)
- **Difetto:** rigidità
- **Le domande:**
 - **Aperte:** chi risponde dà la risposta che vuole (difficile analizzarle)
 - **Chiuse:** chi risponde seleziona una risposta in un insieme limitato

Problemi di progetto del questionario

- Rendere le **domande non ambigue**
- **Aumentare la probabilità che chi deve rispondere lo completi:** questionario corto, remunerare chi compila, invio risposta non costoso
- **Analisi dei dati:** attenti alle scale! quale statistica è adatta?
- Suggerimenti:
 - fare solo domande di cui si vuole veramente conoscere la risposta
 - condurre un **Test Pilota:** in generale il questionario NON si può adattare all'utente, il questionario costa, il test pilota consente di ridurre i pericoli di incomprensione e ambiguità

Le scale adottabili (1)

Lista di controlli

non c'è ordine, zero, distanza,
statistica corretta: frequenza

Can you use the following text editing commands?

DUPLICATE yes no don't know

PASTE yes no don't know

Figure 30.1 A simple checklist.

Scala di valutazione a punti

C'è un ordine, ma non sono
definiti lo zero e la distanza

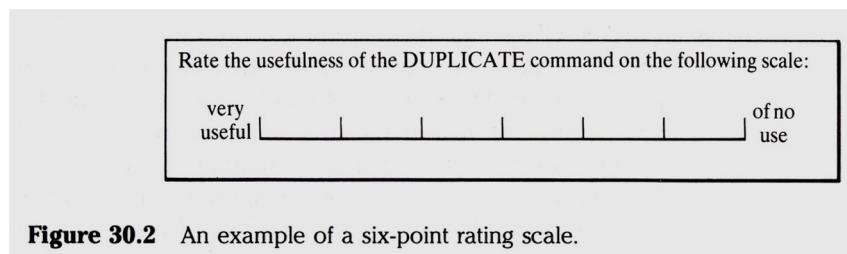


Figure 30.2 An example of a six-point rating scale.

Scala di Likert

Una variante della scala di
valutazione a punti; indaga
accordo/disaccordo con
affermazioni specifiche

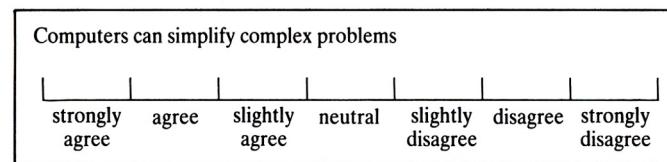


Figure 30.3 An example of a Likert scale.

Le scale adottabili (2)

Differenziale semantico

agli estremi aggettivi bi-polari:
ogni punto dà un peso arbitrario
ad uno dei due

Figure 30.4 A semantic differential example.

Ordinale secondo il rango

ordinare elementi simili rispetto ad una qualità (utilità, difficoltà d'uso)

Place the following commands in order of usefulness: (use a scale of 1 to 4 where 1 is the most useful)							
PASTE	<input type="checkbox"/>	DUPLICATE	<input type="checkbox"/>	GROUP	<input type="checkbox"/>	CLEAR	<input type="checkbox"/>

Figure 30.5 An example of a ranked order questionnaire.

L'analisi dei dati

In generale le scale sono convertite arbitrariamente **a scale numeriche**

strongly disagree	→ -3
disagree	→ -2
slightly disagree	→ -1
neutral	→ 0
slightly agree	→ 1
agree	→ 2
strongly agree	→ 3

e valutate con statistiche per scale numeriche

Quali scale?

- **A tre, cinque o sette elementi?**

- Pochi elementi quando le possibilità sono limitate
- Tanti elementi quando si vuole un giudizio fine da parte degli utenti (e.g., per la UX)

- **Con un numero dispari o pari di elementi?**

- Dipende dagli scopi: vogliamo che gli utenti si esprimano in una direzione precisa?

Un esempio (con difetti)

EUROCHANGE QUESTIONNAIRE

Background information

Departure lounge	Flight	Destination
------------------	--------	-------------

How often do you make European flights?
(please tick one box)

- less than once a month
- 2–4 times a month
- 4–8 times a month
- other (please specify)

Do you normally get your foreign currency at?
(please tick one box)

- own bank
- any convenient bank
- special bureau de change
- airport bank

Please indicate the reasons for using this machine:
(you can tick more than one box)

- did not know where airport bank was
- long queues at airport bank
- last minute need for extra currency
- wanted to try out machine
- shortage of time – had to go directly to departures

The Eurochange machine is a useful addition to airport services:
(please tick one point)

<input type="text"/>				
strongly agree		neutral		strongly disagree

The Eurochange machine should be more widely available at European airports:

<input type="text"/>				
strongly agree		neutral		strongly disagree

Rate the Eurochange machine on the following dimensions:

	extremely	quite	slightly	neutral	slightly	quite	extremely	
predictable	<input type="text"/>	haphazard						
easy	<input type="text"/>	difficult						
simple	<input type="text"/>	complicated						
clear	<input type="text"/>	confusing						
fast	<input type="text"/>	slow						

Figure 30.6 The proposed Eurochange questionnaire.

Questionari generici

- Esistono alcuni questionari con domande che vanno bene per qualunque sistema
- Alcuni sono considerati degli standard, ad esempio:
 - **Questionario SUS (System Usability Scale)** – 10 domande - (Brooke, 1986)
 - Consente il calcolo di un numero che diventa un indicatore di usabilità confrontabile con un valore di riferimento
 - **Questionario PSSUQ (Post-Study System Usability Questionnaire) o CSUQ (Computer System Usability Questionnaire)** – 19 domande in origine (Lewis, 1995), poi ridotte a 16

Questionário SUS

(Brooke, 1986)

© Digital Equipment Corporation, 1986.

1. I think that I would like to use this system frequently
 2. I found the system unnecessarily complex
 3. I thought the system was easy to use
 4. I think that I would need the support of a technical person to be able to use this system
 5. I found the various functions in this system were well integrated
 6. I thought there was too much inconsistency in this system
 7. I would imagine that most people would learn to use this system very quickly
 8. I found the system very cumbersome to use
 9. I felt very confident using the system
 10. I needed to learn a lot of things before I could get going with this system

1	2	3	4	5

1	2	3	4	5

1	2	3	4	5

1	2	3	4	5

1	2	3	4	5	

--	--	--	--	--

1	2	3	4	5

--	--	--	--	--

--	--	--	--	--

--	--	--	--	--

Questionario SUS in italiano

Fornire il grado di accordo con le seguenti affermazioni, dove 1 equivale a "per nulla d'accordo" e 5 a "completamente d'accordo"

1. Penso che mi piacerebbe utilizzare questo sistema frequentemente

1	2	3	4	5
---	---	---	---	---

2. Ho trovato il sistema complesso senza che ce ne fosse bisogno

1	2	3	4	5
---	---	---	---	---

3. Ho trovato il sistema molto semplice da usare

1	2	3	4	5
---	---	---	---	---

4. Penso che avrei bisogno del supporto di una persona già in grado di utilizzare il sistema

1	2	3	4	5
---	---	---	---	---

5. Ho trovato le varie funzionalità del sistema bene integrate

1	2	3	4	5
---	---	---	---	---

6. Ho trovato incoerenze tra le varie funzionalità del sistema

1	2	3	4	5
---	---	---	---	---

7. Penso che la maggior parte delle persone potrebbero imparare ad utilizzare il sistema facilmente

1	2	3	4	5
---	---	---	---	---

8. Ho trovato il sistema molto macchinoso da utilizzare

1	2	3	4	5
---	---	---	---	---

9. Ho avuto molta confidenza con il sistema durante l'uso

1	2	3	4	5
---	---	---	---	---

10. Ho avuto bisogno di imparare molti processi prima di riuscire ad utilizzare al meglio il sistema

1	2	3	4	5
---	---	---	---	---

Versione in italiano per i siti web

Fa parte delle tecniche suggerite nel
protocollo eGLU

(GLU: Gruppo di Lavoro per l'Usabilità -
Ministero per la Pubblica Amministrazione)

	Fortemente in disaccordo				Fortemente d'accordo
1. Penso che mi piacerebbe utilizzare questo sito frequentemente	1	2	3	4	5
2. Ho trovato il sito inutilmente complesso	1	2	3	4	5
3. Ho trovato il sito molto semplice da usare	1	2	3	4	5
4. Penso che avrei bisogno del supporto di una persona già in grado di utilizzare il sito	1	2	3	4	5
5. Ho trovato le varie funzionalità del sito bene integrate	1	2	3	4	5
6. Ho trovato incoerenze tra le varie funzionalità del sito	1	2	3	4	5
7. Penso che la maggior parte delle persone possano imparare ad utilizzare il sito facilmente	1	2	3	4	5
8. Ho trovato il sito molto difficile da utilizzare	1	2	3	4	5
9. Mi sono sentito a mio agio nell'utilizzare il sito	1	2	3	4	5
10. Ho avuto bisogno di imparare molti processi prima di riuscire ad utilizzare al meglio il sito	1	2	3	4	5

Metodo di Misura (Brooke, 1986)

- SUS arriva a un **numero (score)** che rappresenta **una misura globale dell'usabilità** del sistema
- Per calcolare il SUS score, prima **si sommano i contributi** per ogni risposta
- Tali contributi vanno da **0 a 4**
- Per le domande **1, 3, 5, 7, 9** il contributo è la **posizione** della risposta sulla scala **meno 1**
- Per le domande **2, 4, 6, 8, 10** il contributo è **5 meno la posizione** della risposta sulla scala
- Si **moltiplica** la somma dei punti così ottenuti **per 2.5** e si ottiene il valore globale del SUS
- I SUS score variano **da 0 a 100**
- Esperimenti hanno derivato che un punteggio **superiore a 68** può essere considerato superiore alla media

1. I think that I would like to use this system frequently
2. I found the system unnecessarily complex
3. I thought the system was easy to use
4. I think that I would need the support of a technical person to be able to use this system
5. I found the various functions in this system were well integrated
6. I thought there was too much inconsistency in this system
7. I would imagine that most people would learn to use this system very quickly
8. I found the system very cumbersome to use
9. I felt very confident using the system
10. I needed to learn a lot of things before I could get going with this system

Strongly disagree						Strongly agree
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/> X	4	
1	2	3	4	5		
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/> X	<input type="checkbox"/>	1	
1	2	3	4	5		
<input type="checkbox"/>	<input checked="" type="checkbox"/> X	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1	
1	2	3	4	5		
<input checked="" type="checkbox"/> X	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	4	
1	2	3	4	5		
<input type="checkbox"/>	<input checked="" type="checkbox"/> X	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1	
1	2	3	4	5		
<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/> X	<input type="checkbox"/>	<input type="checkbox"/>	2	
1	2	3	4	5		
<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/> X	<input type="checkbox"/>	<input type="checkbox"/>	1	
1	2	3	4	5		
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/> X	<input type="checkbox"/>	1	
1	2	3	4	5		
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/> X	4	
1	2	3	4	5		
<input type="checkbox"/>	<input checked="" type="checkbox"/> X	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	3	
1	2	3	4	5		

Total score = 22

SUS score = $22 * 2.5 = 55$

PSSUQ

The Post-Study Usability Questionnaire
Version 3

	The Post-Study Usability Questionnaire Version 3	Strongly agree	Strongly disagree							NA
			1	2	3	4	5	6	7	
1	Overall, I am satisfied with how easy it is to use this system.		○	○	○	○	○	○	○	○
2	It was simple to use this system.		○	○	○	○	○	○	○	○
3	I was able to complete the tasks and scenarios quickly using this system.		○	○	○	○	○	○	○	○
4	I felt comfortable using this system.		○	○	○	○	○	○	○	○
5	It was easy to learn to use this system.		○	○	○	○	○	○	○	○
6	I believe I could become productive quickly using this system.		○	○	○	○	○	○	○	○
7	The system gave error messages that clearly told me how to fix problems.		○	○	○	○	○	○	○	○
8	Whenever I made a mistake using the system, I could recover easily and quickly.		○	○	○	○	○	○	○	○
9	The information (such as online help, on-screen messages and other documentation) provided with this system was clear.		○	○	○	○	○	○	○	○
10	It was easy to find the information I needed.		○	○	○	○	○	○	○	○
11	The information was effective in helping me complete the tasks and scenarios.		○	○	○	○	○	○	○	○
12	The organization of information on the system screens was clear.		○	○	○	○	○	○	○	○
13	The interface* of this system was pleasant.		○	○	○	○	○	○	○	○
14	I liked using the interface of this system.		○	○	○	○	○	○	○	○
15	This system has all the functions and capabilities I expect it to have.		○	○	○	○	○	○	○	○
16	Overall, I am satisfied with this system.		○	○	○	○	○	○	○	○

*The "interface" includes those items that you use to interact with the system. For example, some components of the interface are the keyboard, the mouse, the microphone, and the screens (including their graphics and language).

CSUQ

The Computer System Usability Questionnaire
Version 3

	Strongly agree								Strongly disagree
	1	2	3	4	5	6	7	NA	

1	Overall, I am satisfied with how easy it is to use this system.	<input type="radio"/>							
2	It is simple to use this system.	<input type="radio"/>							
3	I am able to complete my work quickly using this system.	<input type="radio"/>							
4	I feel comfortable using this system.	<input type="radio"/>							
5	It was easy to learn to use this system.	<input type="radio"/>							
6	I believe I became productive quickly using this system.	<input type="radio"/>							
7	The system gives error messages that clearly tell me how to fix problems.	<input type="radio"/>							
8	Whenever I make a mistake using the system, I recover easily and quickly.	<input type="radio"/>							
9	The information (such as online help, on-screen messages, and other documentation) provided with this system is clear.	<input type="radio"/>							
10	It is easy to find the information I needed.	<input type="radio"/>							
11	The information provided with the system is effective in helping me complete my work.	<input type="radio"/>							
12	The organization of information on the system screens is clear.	<input type="radio"/>							
13	The interface* of this system is pleasant.	<input type="radio"/>							
14	I like using the interface of this system.	<input type="radio"/>							
15	This system has all the functions and capabilities I expect it to have.	<input type="radio"/>							
16	Overall, I am satisfied with this system.	<input type="radio"/>							

*The "interface" includes those items that you use to interact with the system. For example, some components of the interface are the keyboard, the mouse, the microphone, and the screens (including their graphics and language).

Come si calcola il risultato?

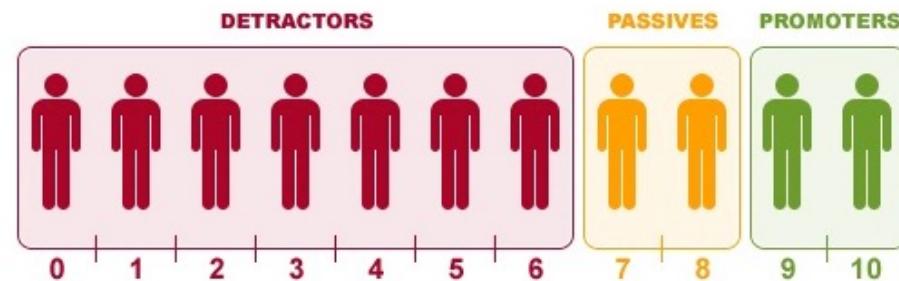
- Gli elementi producono quattro punteggi:
 - **Overall** (punteggio complessivo) media fra le risposte a tutti gli elementi del questionario (voci da 1 a 16)
 - **SysQual** (qualità del sistema): media per gli elementi da 1 a 6
 - **InfoQual** (qualità delle informazioni): media per gli elementi da 7 a 12
 - **IntQual** (qualità dell'interfaccia): media per gli elementi da 13 a 15
- E' suscettibile di **bias di acquiescenza** (o bias di accordo): la tendenza degli intervistati ad essere d'accordo con le affermazioni della ricerca

Net Promoter Score

- 1 sola domanda (non un questionario vero e proprio)
- Per misurare la “fedeltà del cliente” (soddisfazione dell’utente)

“Attribuendo un valore compreso tra 0 e 10, quanto consiglieresti l’azienda/prodotto/sistema X a un amico/parente/collega?”, in cui 0 è il valore più negativo e 10 il più positivo della scala di soddisfazione

Il calcolo del NPS

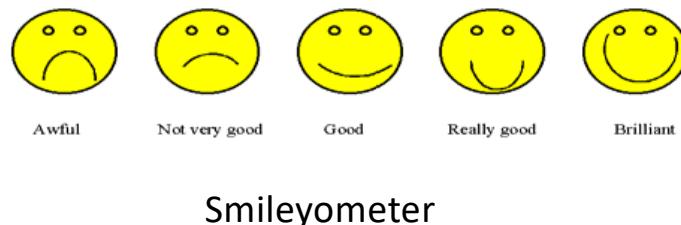


$$\text{Net Promoter Score} = \% \text{ Promoters} - \% \text{ Detractors}$$

- Si suddividono le risposte in 3 categorie:
 - **I promotori:** coloro che rispondono 9 o 10
 - **I passivi:** coloro che rispondono 7 o 8
 - **I detrattori:** coloro che rispondono da 0 a 6
- Si **sottrae la percentuale di detrattori alla percentuale di promotori** e il risultato viene espresso come numero assoluto compreso **tra -100 e +100**
- L'ideale sarebbe ottenere 100, sarebbe bene **non scendere sotto 50**

Raccogliere opinioni dai bambini

- Se gli utenti sono **bambini** i questionari sono semplificati, così come la loro somministrazione
- Esempi:



Would you like to do it again?

	Yes	Maybe	No
clock	✓		
drive		✓	

Read, J. C. & MacFarlane, S. (2006). Using the fun toolkit and other survey methods to gather opinions in child computer interaction. In *Proceedings of the 2006 Conference on Interaction Design and Children*, IDC '06.

Name of child..... Age..... Boy / Girl

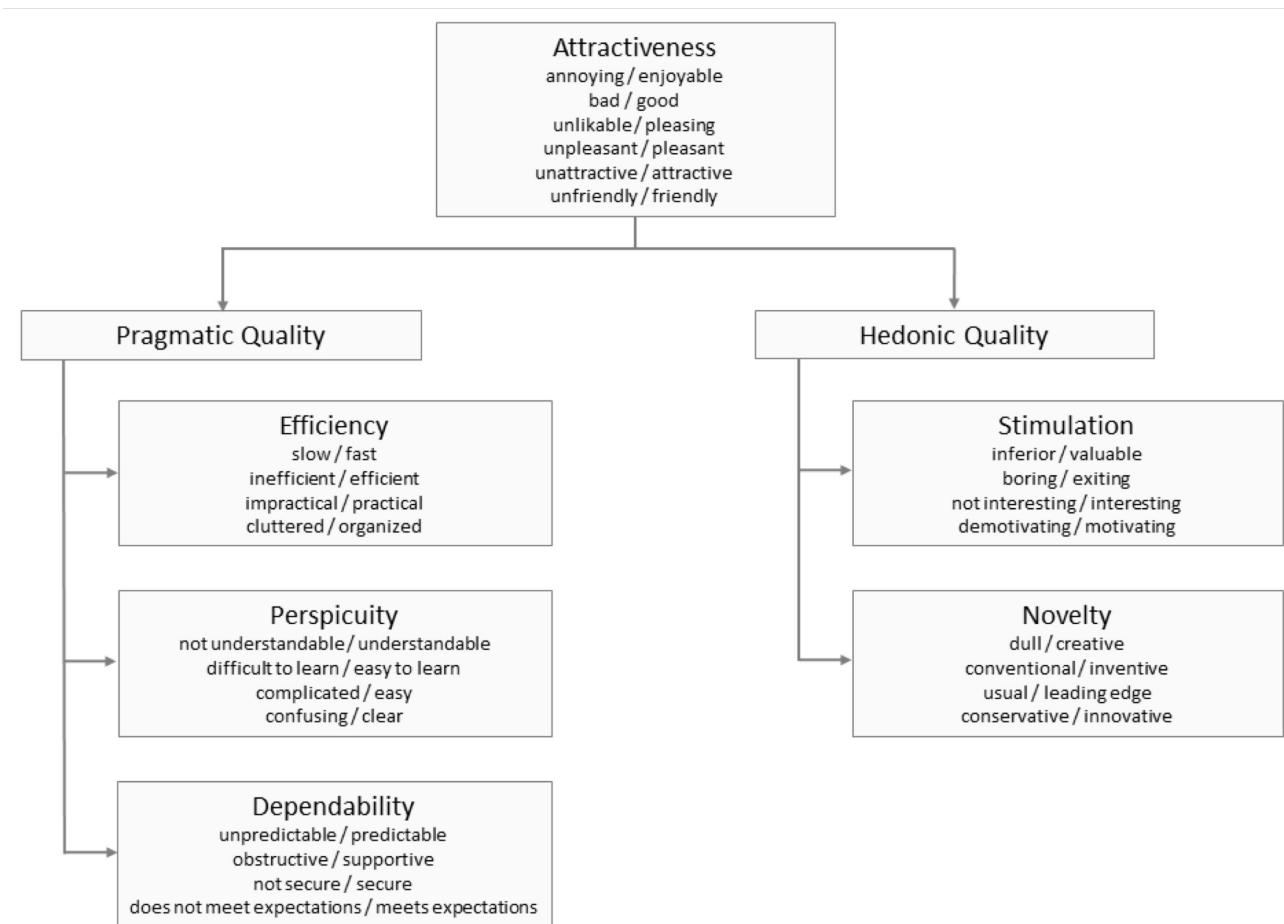
Again – Again table

Non solo usabilità...

- **UEQ (User Experience Questionnaire)**: <http://www.ueq-online.org>
- Questionari nelle varie lingue, tool Excel per l'analisi dei risultati (anche per scopi di confronto)
- Un questionario di **26 item (differenziale semantico)**
- Esempio di item:

attractive o o o o o o unattractive
- Le risposte sono mappate sulla scala {-3, -2, -1, 0, +1, +2, +3}

Classificazione dei 26 item



UEQ: La versione italiana

È importante che i partecipanti compilino il questionario nella loro lingua nativa

Richiede 3-5 minuti per la compilazione

Ora inserisca per favore la Sua valutazione del prodotto. Per ogni coppia di aggettivi è possibile dare una sola risposta.

	1	2	3	4	5	6	7		
fastidioso	<input type="radio"/>	piacevole	1						
incomprensibile	<input type="radio"/>	comprendibile	2						
creativo	<input type="radio"/>	privo di fantasia	3						
facile da apprendere	<input type="radio"/>	difficile da apprendere	4						
di grande valore	<input type="radio"/>	di poco valore	5						
noioso	<input type="radio"/>	appassionante	6						
non interessante	<input type="radio"/>	interessante	7						
imprevedibile	<input type="radio"/>	prevedibile	8						
veloce	<input type="radio"/>	lento	9						
originale	<input type="radio"/>	convenzionale	10						
ostrettivo	<input type="radio"/>	di supporto	11						
buono	<input type="radio"/>	scarso	12						
complicato	<input type="radio"/>	facile	13						
repellente	<input type="radio"/>	attraente	14						
usuale	<input type="radio"/>	moderno	15						
sgradevole	<input type="radio"/>	gradevole	16						
sicuro	<input type="radio"/>	insicuro	17						
attivante	<input type="radio"/>	soporifero	18						
conforme alle aspettative	<input type="radio"/>	non conforme alle aspettative	19						
inefficiente	<input type="radio"/>	efficiente	20						
chiaro	<input type="radio"/>	confuso	21						
non pragmatico	<input type="radio"/>	pragmatico	22						
ordinato	<input type="radio"/>	sovraffatto	23						
invitante	<input type="radio"/>	non invitante	24						
congeniale	<input type="radio"/>	ostile	25						
conservativo	<input type="radio"/>	innovativo	26						

UEQ Benchmark

- Contiene i dati di 246 valutazioni di prodotti (9905 partecipanti in totale)

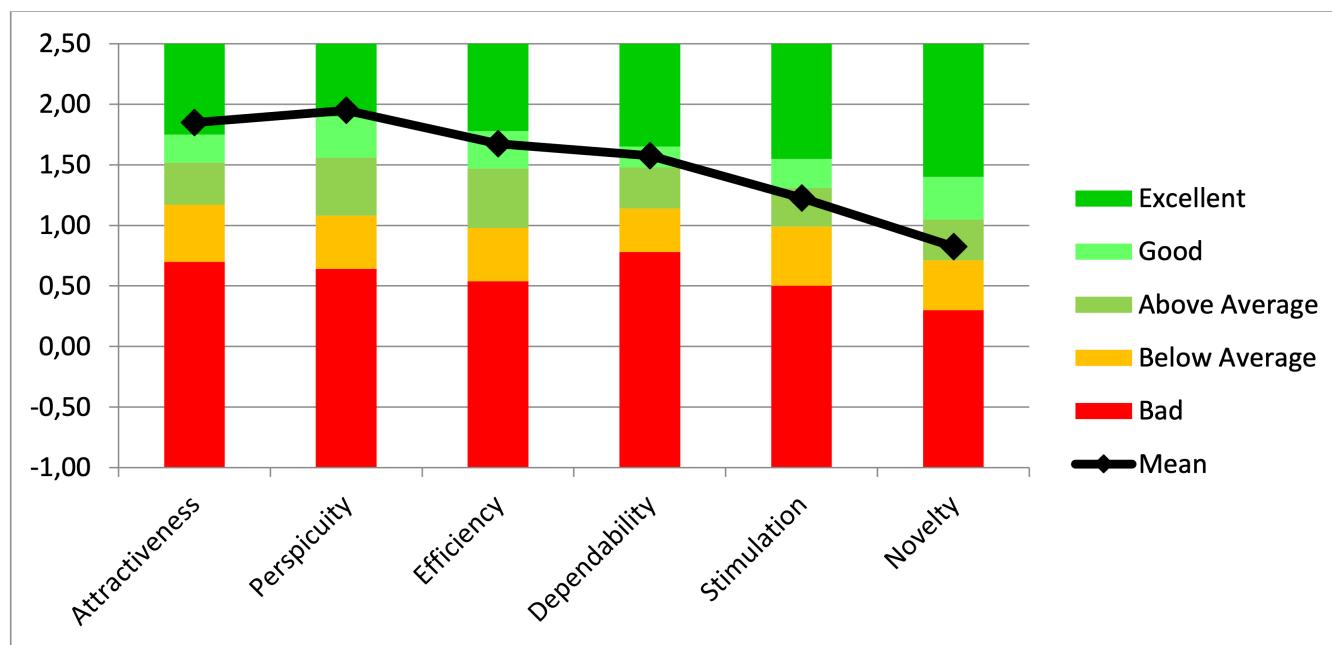


Figure 4: Benchmark graph for a hypothetical product.

Schrepp, M.; Hinderks, A. & Thomaschewski, J. (2017). Construction of a benchmark for the User Experience Questionnaire (UEQ). International Journal of Interactive Multimedia and Artificial Intelligence, 4(4), 40-44.

Versione “short”: UEQ-S

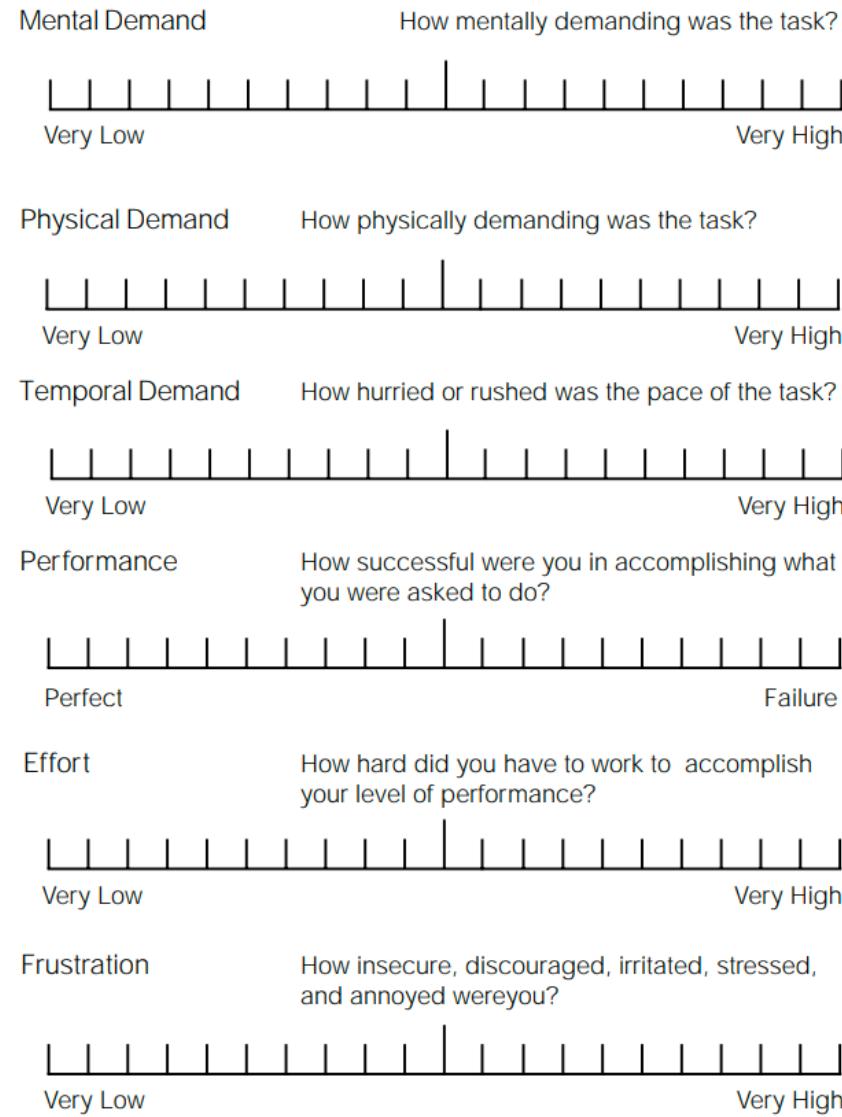
ostruttiva	○ ○ ○ ○ ○ ○	di supporto	Pragmatic Quality
complicato	○ ○ ○ ○ ○ ○	facile	
inefficiente	○ ○ ○ ○ ○ ○	efficiente	
confuso	○ ○ ○ ○ ○ ○	chiaro	
noioso	○ ○ ○ ○ ○ ○	appassionante	Hedonic Quality
non interessante	○ ○ ○ ○ ○ ○	interessante	
convenzionale	○ ○ ○ ○ ○ ○	originale	
usuale	○ ○ ○ ○ ○ ○	moderno	

- 3 scenari in cui risulta utile:
 - L’utente esce da un negozio/servizio su web
 - Come aggiunta a un questionario di “customer experience”
 - Quando il partecipante deve valutare tanti prodotti o varianti

NASA-TLX

(Hart and Staveland, 1988)

- **NASA Task Load Index:** stima il **carico di lavoro percepito** dagli utenti nell'esecuzione di un compito
- Questionario con 6 item che misurano il carico di lavoro attraverso **6 dimensioni soggettive:**
 - Mental Demand
 - Physical Demand
 - Temporal Demand
 - Performance
 - Effort
 - Frustration



Il calcolo dell'indicatore NASA-TLX

- Nella versione originale vengono fatte domande aggiuntive agli utenti per assegnare **pesi diversi alle dimensioni** e poi viene fatto il calcolo
- Nella versione "**Raw TLX**"
 - Ogni dimensione riceve un punteggio nell'intervallo 5-100, poiché è valutata su una scala a 20 punti, ciascuno di valore 5
 - Il carico di lavoro riferito a un utente è il punteggio medio sulle 6 dimensioni
 - Il carico di lavoro riferito a un sistema si ottiene facendo la media di tutti i valori dei carichi di lavoro degli utenti

App su Apple Store e
Google Play

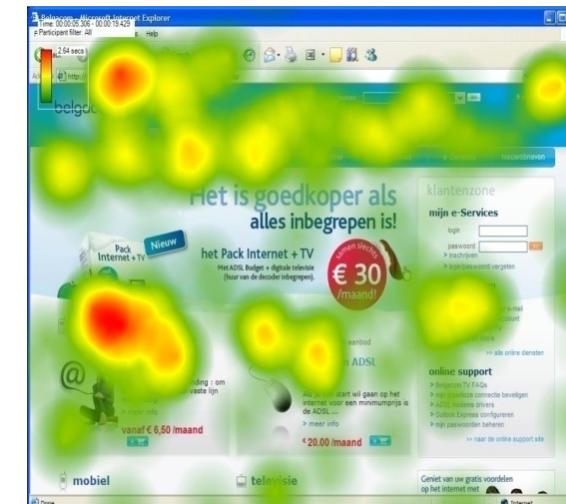


In generale sui questionari online

- Applicazioni web facilitano la preparazione delle domande e la codifica delle risposte
- Esempi: Google Form, SurveyMonkey, etc.
- **Vantaggi:**
 - Basso costo
 - Efficienza e velocità
 - Possibilità di raggiungere vasti campioni, o campioni di partecipanti particolari
 - Possibilità di ricerche che indagano aspetti di tipo culturale o interculturale (es: Europei vs. Giapponesi vs. Americani, etc.)
- **Svantaggi:**
 - Polarizzazione nella scelta del campione
 - Difficoltà nel controllare le caratteristiche e la serietà di chi compila
 - Difficoltà di monitoraggio del contesto in cui avviene la compilazione

Monitoraggio utenti (Logging)

- Si tratta di osservare, mediante **programmi**, gli utenti mentre interagiscono con un sistema software
- Si avranno statistiche sulla **frequenza** con cui l'utente ha usato una certa funzionalità o con cui si sono verificati certi eventi di interesse (esempio, messaggi di errore)
- Spesso usato **nei siti web (web analytics)**, per capire i percorsi che fanno gli utenti, il tempo speso sulle pagine, chi ha visitato il sito, da quale sito i visitatori provengono, dove i visitatori lasciano il sito, etc.
- Anche con uso di **eyetracker**



Daniela Fogli - Interazione Persona-Calcolatore

Monitoraggio di tipo avanzato: “mobile eye tracking”



Fig. 5. Advanced mobile EAID makes use of a scene camera and eye tracking to monitor both the action and the attention of the user. Today's mobile eye-tracking systems provide unobtrusive means to observe user behaviors in many interaction contexts, from reading or operating a desktop computer, to car driving or sports.

[T. Pfeiffer, J. Hellmers, E. Schon, J. Thomaschewski,
Empowering User Interfaces for Industrie 4.0,
Proceedings of the IEEE, 2016]

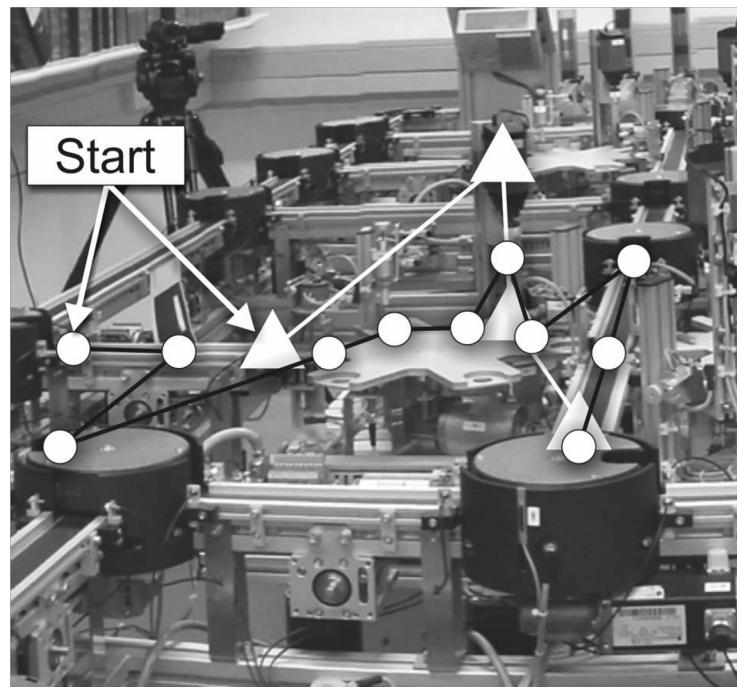


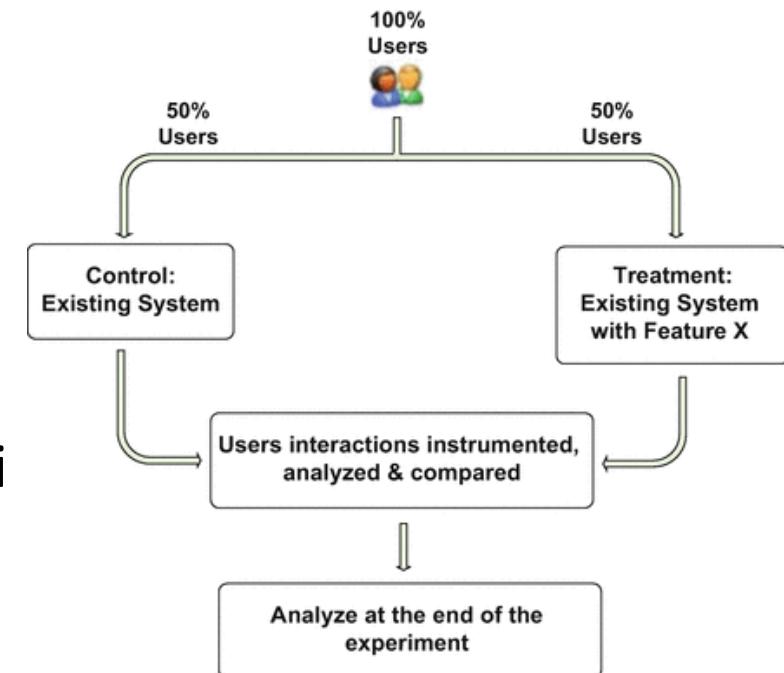
Fig. 7. The sequence with which the machine has been observed by a novice operator (circles) and an expert (triangles). Each symbol represents a fixation on that particular location of the machine and the size of the symbol represents the duration of that fixation. The two sequences differ significantly and allow the observer to assess the depth of understanding of the machine.

Monitoraggio utenti: a che livello

- Con strumenti di **basso livello** (battitura tasti o mouse click *time-stamped*): facile da fare ma produce massa risultati costosa da esaminare
- A livello di **gestori di interfaccia grafica**: cattura eventi ‘interpretati’ (e.g. selezione dell’icona → seleziona l’archivio pazienti)
- Con **agenti osservatori**: moduli software che individuano pattern di attività di alto livello

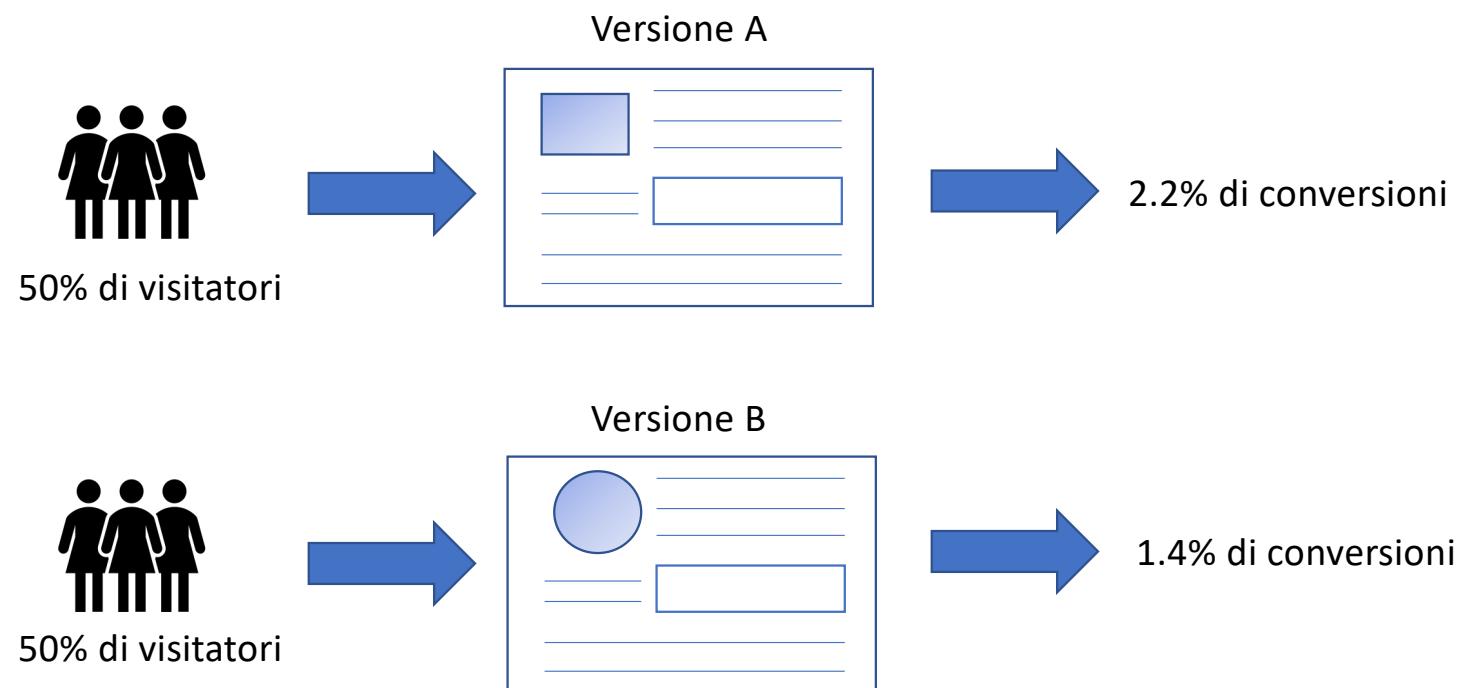
A/B Testing

- Versione online dell'esperimento controllato
- Coinvolge un numero molto elevato di utenti in modalità **between-subjects**
- Gli utenti possono **non sapere** di essere i partecipanti a un esperimento



Kohavi R., Longbotham R. (2017) Online Controlled Experiments and A/B Testing. In: Sammut C., Webb G.I. (eds) Encyclopedia of Machine Learning and Data Mining. Springer, Boston, MA.

Spesso usato per valutare il tasso di conversione



Pro e contro del monitoraggio

- **Vantaggi:**

- Poco intrusivo (ma dipende dallo strumento)
- Mette a disposizione grandi quantità di dati riesaminabili
- Non richiede presenza valutatore

- **Svantaggi:**

- Problemi di privacy
- Servono strumenti di visualizzazione ed esplorazione dati

Osservazione, monitoraggio, raccolta opinioni: i documenti

- **Osservazione** (dell'utente in attività): deve risultare in documenti che descrivono l'attività e in documenti che sintetizzano i risultati
- **Questionari ed interviste** (raccolta di opinioni dell'utente): risultano in archivi di rapporti con osservazioni emerse da interviste e da questionari compilati
- **Monitoraggio** (dell'utente che svolge una attività): risulta in data log e in documenti che sintetizzano i risultati

Problemi comuni:

progettazione accurata, esecuzione e analisi dati metodica,
necessità di rassicurare, retribuire e motivare gli utenti

Confronto fra i metodi di valutazione

Obiettivo

- Capire pro e contro dei metodi visti per
 - selezionare i metodi adatti al caso in esame e pianificare la strategia di valutazione
 - criticare costruttivamente piani e studi fatti da altri

Fattori per il confronto

- **Scopo della valutazione**

- Il sistema è buono, è adatto, presenta problemi?
- Confrontare progetti alternativi: qual è il migliore? (rispetto a cosa?)
- Capire il mondo reale: come funziona nella situazione reale?

- **Stadio dello sviluppo**

- Rispetto al ciclo di vita del software interattivo

- **Tipi di dati ottenibili**

- **Coinvolgimento degli utenti**

- **Risorse disponibili**

- Sono disponibili gli strumenti? Sono disponibili gli esperti?
- C'è tempo a sufficienza? Sia per la valutazione sia per l'analisi dei dati?
- Ci sono gli utenti e sono disponibili? Per quanto? Quanti?
- Budget?

Scopo della valutazione

Metodo	Scopo della valutazione
Predittivi	Individuare problemi di usabilità, ingegnerizzare rispetto a un obiettivo (KLM)
Esperimento	Confrontare sistemi/prodotti diversi, ingegnerizzare rispetto a un obiettivo
Osservazione	Comprendere il mondo reale
Raccogliere le opinioni degli utenti	Comprendere il mondo reale, confrontare sistemi, individuare problemi di usabilità
Monitoraggio	Comprendere il mondo reale, confrontare sistemi

Stadio dello sviluppo

Metodo	Stadio dello sviluppo
Predittivi	Soltamente nelle fasi iniziali dello sviluppo
Esperimento	Ci deve essere un prototipo funzionante
Osservazione	Qualsiasi stadio dello sviluppo
Raccogliere le opinioni degli utenti	Qualsiasi stadio dello sviluppo
Monitoraggio	Di solito su un sistema in uso

Tipi di dati ottenibili

Metodo	Tipi di dati
Predittivi	Prevalentemente qualitativi (quantitativi con KLM)
Esperimento	Quantitativi
Osservazione	Generalmente qualitativi, ma uno stream video può generare dati sia quantitativi che qualitativi
Raccogliere le opinioni degli utenti	I questionari producono dati quantitativi, le interviste e i focus group dati qualitativi
Monitoraggio	L'analisi dei log genera di solito dati quantitativi

Coinvolgimento degli utenti

Metodo	Partecipazione degli utenti	Controllo dell'utente sul compito	Controllo dell'utente sulla valutazione
Predittivi	No	NA	NA
Eperimento	Sì	Variabile	Nessuno
Osservazione	Sì	Sì	Variabile
Raccogliere le opinioni degli utenti	Sì	NA	Nessuno
Monitoraggio	Sì	Sì	Nessuno

Altri fattori per il confronto

- **Criteri tecnici** relativi ai dettagli dell'uso delle tecniche di valutazione
 - **validità interna:** quanto un metodo misura ciò che vogliamo valutare
 - **validità esterna:** quanto i risultati sono generalizzabili
 - **validità “ecologica”:** come il metodo di valutazione influenza i risultati ottenuti (intrusività del metodo, il laboratorio è spesso poco realistico, ...)
 - **polarizzazione (bias):** quanto i risultati possono essere distorti (bias degli esperti di usabilità in valutazioni predittive, degli utenti nelle interviste, ...)
 - **completezza** dei dati raccolti (rispetto ai compiti provati, schermate analizzate, ...)

I metodi a confronto

Metodo	Vantaggi	Svantaggi
Predittivi	Non richiedono sistema funzionante Trovano problemi di usabilità specifici e le loro ragioni	Alcuni hanno un focus limitato Non coinvolgono gli utenti, non trovano “sorprese”
Esperimenti	Forniscono misure di usabilità I risultati quantitativi possono servire per confrontare sistemi	Sono costosi Se di tipo controllato possono non trovare problemi di usabilità specifici
Osservazione	Ampiamente applicabili, scoprono le difficoltà degli utenti Poco costosi	Possono influire sul comportamento dell’utente Alcune forme sono innaturali per gli utenti (carico cognitivo)
Monitoraggio	Trovano dati interessanti (tipo funzionalità usate e non usate)	Occorrono programmi per analizzare grandi moli di dati, problemi di privacy
Opinioni degli utenti	Dicono preferenze soggettive	Ci può essere un basso tasso di risposta (questionari) A volte i dati sono difficili da interpretare (interviste)