

Francesco Di Nocera

Usability Evaluation 2.0

Una descrizione (s)oggettiva dell'usabilità



Titolo originale dell'opera:

Usability Evaluation 2.0: Una descrizione (s)oggettiva dell'usabilità

di Francesco Di Nocera

Copyright © 2013 Francesco Di Nocera

Questo volume è distribuito gratuitamente per volontà dell'autore. TUTTI I DIRITTI RISERVATI, ad eccezione del questionario Us.E. 2.0 rilasciato con licenza CC BY-ND.

Edizioni

Ergoproject Srl

Via Antonio Pacinotti, 63 – 00146 Roma

Tel./Fax +39 06 45425466

info@ergoproject.it

www.ergoproject.it

I edizione – Copyright © 2013 Francesco Di Nocera

ISBN 978-88-90952-70-8

Progetto grafico di Maria Luce Lupetti

Francesco Di Nocera

Usability Evaluation 2.0: Una descrizione (s)oggettiva dell'usabilità

Indice

Prefazione di Maurizio Boscarol	6
Premessa	10
Avvertenza	13
1. Che cos'è l'usabilità	15
1.1. Il concetto di interfaccia	16
1.2. Usabilità dei siti Internet	19
2. Misurare l'usabilità	22
2.1 Stima dell'usabilità senza il coinvolgimento degli utenti	25
2.2 Stima dell'usabilità con coinvolgimento degli utenti	26
2.3 L'uso dei questionari nella valutazione dell'usabilità	30
3. Il Progetto Us.E.	33
3.1 Timeline	33
3.2 Dimensioni	34
3.2.1 Maneggevolezza	35
3.2.2 Soddisfazione	37
3.2.3 Attrattiva	38
3.3 Anagrafica e opinioni degli utenti	38
3.4 Licenza d'uso	39
3.5 Controindicazioni	40

4.	Proprietà psicometriche	41
4.1	Campione	41
4.2	Struttura Fattoriale e Attendibilità	43
4.3	Validità	46
5.	Somministrazione, scoring e interpretazione dei punteggi	48
5.1	Somministrazione	49
5.2	Numerosità minima	50
5.3	Norme di taratura	52
5.4	Scoring, standardizzazione e interpretazione dei punteggi	55
5.5	Sito di supporto al questionario	60
	Appendice 1: Come scrivere un report di valutazione	64
	Appendice 2: Us.E. 2.0 - English Version	70
	Riferimenti bibliografici	72

Prefazione di Maurizio Boscarol

Conta più l'usabilità o l'usabilità percepita? Dietro questa domanda si nascondono insidie che ogni specialista che lavori nel campo dell'usabilità conosce o dovrebbe conoscere. Per come è definita nella ISO 9241, l'usabilità incorpora già almeno una componente soggettiva, di "percezione" della stessa: viene chiamata "soddisfazione", laddove "efficacia" ed "efficienza" sono invece le componenti prestazionali, oggettive, che tutti siamo chiamati a osservare e misurare nei test con utenti. Proprio dietro questo equivoco fra soddisfazione e usabilità percepita si gioca la partita dei diversi questionari tradizionalmente impiegati per valutare rapidamente gli aspetti soggettivi dell'esperienza d'uso. Alcuni di questi strumenti, come sottolinea Francesco Di Nocera in questo manuale, sono centrati proprio sul concetto di "soddisfazione". Si tratta, però, di un termine che riguarda di più il marketing e che in ogni caso, per quanto sappiamo essere correlato in qualche modo con un costrutto multifattoriale come l'usabilità, non è esattamente quello che vogliamo valutare con un questionario di usabilità.

Ma cosa vogliamo esattamente valutare con un questionario di usabilità? Come sappiamo che è qualcosa di diverso dalla semplice soddisfazione? Quale studio, quale articolo, quale trattato tecnico-scientifico ci spiega esattamente cosa sia l'usabilità percepita? Nonostante non sia disponibile una spiegazione esaustiva, questo piccolo manuale d'uso del questionario Usability Evaluation 2.0 chiarisce in parte queste domande, non solo mettendo a disposizione della comunità scientifica e professionale uno strumento standardizzato, ma anche ricavando, attraverso la storia del suo sviluppo, informazioni preziose sulla relazione tra utenti e siti. Ciò che differenzia questo strumento da altri (solo in apparenza simili)

è proprio il modo in cui è stato pensato fin dall'origine: con un approccio bottom-up, basandosi sulle affermazioni di utenti che usavano dei siti e senza assunzioni teoriche da confermare o da confutare. Questo ha fatto sì che tra le svariate decine di item che componevano le prime formulazioni di Us.E. sia stato possibile lavorare progressivamente per mantenere solo quelli che meglio si aggregavano per "concentrati di senso" e che hanno formato, alla fine, un costrutto originale e inedito, articolato su tre dimensioni, che nel questionario sono valutate da altrettante sotto-scale che ci dicono qualcosa di molto più preciso sull'usabilità percepita di quanto non facciano altri questionari a nostra disposizione.

Le tre dimensioni sono la Maneggevolezza, la Soddisfazione (toh! Allora è solo una delle componenti soggettive che vogliamo valutare, non l'unica, né la principale...) e l'Attrattiva. Le etichette di Maneggevolezza, Soddisfazione e Attrattiva, come tutte le etichette, potrebbero generare malintesi o darci un'idea parziale dei concetti che sottendono. Leggendo gli item che compongono le sotto-scale (e leggendo questo manuale), possiamo capire meglio cosa si intenda con queste tre dimensioni e renderci conto che l'usabilità dei siti, per come è percepita dagli utenti, comprende almeno questo:

1. La capacità di muoverci nel sito, di operare con esso. Sia a livello di utilizzo delle singole pagine, che di relazione fra di esse. Una qualità strutturale, che ha a che fare sia con il layout, con la disposizione degli elementi in pagina, sia con quella che oggi si chiama architettura dell'informazione e che impatta sulla navigazione.

2. La soddisfazione nell'uso dei contenuti del sito. Che è cosa diversa sia dalla soddisfazione generale sia dalla gradevolezza: coincide con una corretta relazione fra il contenuto o il servizio

progettato per quel sito e il pubblico che quel sito incontra.

3. La piacevolezza, che è in questo caso soprattutto visiva; l'idea che il sito sia gradevole, e che è - ormai diverse ricerche lo suggeriscono - strettamente correlata con la disponibilità di un utente a investire tempo visitando un sito, anche superando eventuali difficoltà legate ai punti 1 e 2.

Uno dei meriti del questionario qui presentato è proprio quello di farci capire meglio cosa sia importante per l'utente di un sito. Us.E. ci fornisce una specie di "costrutto-ombra", se mi passate l'espressione, dell'usabilità percepita. Di conseguenza, ci consente di pensare anche alle fasi della progettazione ragionando separatamente sulla creazione di siti maneggevoli, di contenuti adeguati, e di pagine belle da vedere. Sono tre diversi aspetti, ma sono i principali, quelli più centrali nella progettazione web. Il monitoraggio di un sito attraverso Us.E. 2.0, peraltro gratuito e assistito da un servizio web, ci offre preziose informazioni sulla nostra progettazione e su dove indirizzare i nostri sforzi. Informazioni che nessun altro questionario fra quelli disponibili ci offre con tanta chiarezza.

Il fatto che sia gratuito, il fatto che si possa avere un'elaborazione automatica dei risultati, il fatto che sia un lavoro interamente documentato e aperto, costituiscono ulteriori benefit e garanzie offerti da un prodotto che è unico nella comunità tecnico-scientifica. Date queste sue caratteristiche, nei gruppi di lavoro del progetto GLU 2.0 del Ministero della Funzione Pubblica (che ho avuto l'onore di coordinare assieme a Simone Borsci) abbiamo scelto di indicare Us.E. come strumento di monitoraggio costante dei siti delle Pubbliche Amministrazioni, anche a prescindere dai test con gli utenti: anzi, come perfetto strumento ponte fra un test e l'altro.

Chiudo questa prefazione sottolineando quello che considero l'unico grande "difetto" dell'altrimenti ottimo lavoro di Di Nocera e del suo gruppo: Us.E. nasce con il peccato originale di essere in italiano. Dobbiamo purtroppo constatare che il mondo dell'usabilità è infatti "English-centered", più che "user-centered". Lavori in qualche caso scadenti, o di qualità opinabile, o magari semplicemente non verificabile, assurgono al rango di strumenti centrali, impiegati in tutto il mondo in virtù della semplice appartenenza alla lingua franca per eccellenza nella nostra comunità di pratiche.

In calce al manuale viene proposta una versione in inglese del questionario, aperta al necessario lavoro di ri-standardizzazione che la renderebbe una vera e propria forma equivalente dell'originale lavoro italiano. Già in questa forma, comunque, ho il sospetto che la versione inglese costituisca un lavoro quanto meno equivalente ad altri questionari più blasonati.

Dal momento che far diventare l'italiano la lingua franca dello user-centered design è un obiettivo non immediatamente plausibile (anche se per competenze e ampiezza culturale se ne sentirebbe il bisogno), non ci resta che sostenere e incoraggiare il lavoro necessario per la taratura della versione inglese di Us.E. 2.0. Nel frattempo possiamo goderci uno strumento con cui arricchire il nostro bagaglio professionale che, nel mondo dell'usabilità, è davvero unico.

Maurizio Boscarol

usabile.it

Premessa

Avere una presenza on-line costituisce oggi un importante indicatore di credibilità e successo per le organizzazioni. Essere privi di un sito Internet rappresenta uno svantaggio competitivo, ma allo stesso tempo una visibilità on-line di scarsa qualità può danneggiare aziende ed enti che vogliono offrire informazioni, servizi o prodotti ai propri clienti/utenti attraverso Internet. Non basta dunque essere meramente presenti, occorre anche curare questa presenza. I siti Internet (ma il discorso coinvolge ormai da tempo anche i social network) non sono più dei semplici avamposti usati per presidiare un canale, ma rappresentano in molti casi il luogo privilegiato, se non unico, in cui avviene l'interazione tra persone, organizzazioni e prodotti.

Internet è diventato il principale strumento di promozione e vendita. Ne consegue che un qualunque progetto per portare in rete contenuti e servizi deve necessariamente partire da un design strategico che consenta di definire la struttura del futuro sito e faciliti gli utenti nel raggiungimento dei loro obiettivi.

Diverse organizzazioni hanno compreso da tempo questa necessità e si affidano a professionisti che -affiancando sviluppatori ed esperti di marketing- concepiscono, progettano e valutano specificatamente l'interfaccia utente allo scopo di garantire la migliore esperienza di interazione (User Experience o UX).

Ciononostante, il cattivo design è ancora molto diffuso e non è difficile imbattersi in siti Internet che non avendo previsto una figura dedicata agli aspetti di interazione, risultano mal progettati e incuranti delle esigenze degli utenti. Spesso, è lo stesso sviluppatore che opera le scelte strategiche relative

l'interfaccia, nel migliore dei casi ricorrendo superficialmente a linee guida per la progettazione. Tuttavia, l'adesione alle linee guida non è sufficiente: la qualità dell'interazione con un sito Internet non è un'informazione cristallizzata e immutabile nel tempo. In un contesto tecnologico che ci espone a mutamenti continui, il rischio di produrre lunghe liste di buone pratiche che invecchiano rapidamente è sempre in agguato e la rigida applicazione di "ricette" può far più danno che portare benefici.

Le linee guida sono ovviamente molto attraenti per la loro parvenza di "leggi" alle quali attenersi e, in molti casi, costituiscono un utile vincolo per evitare errori grossolani, ma un esperto di usabilità (ricercatore o consulente che sia) sa bene che non vi è una risposta univoca e definitiva alla domanda "cosa occorre fare per rendere usabile questo servizio?". Piuttosto, dispone di un insieme di metodologie e conoscenze da applicare in maniera creativa (non "artistica" e "opinabile", ma creativa nel senso che comporta la creazione di una soluzione), misurabile e documentabile. Trattare il problema dell'usabilità solo in modo normativo non è dunque una strategia efficace. Non si tratta di un'attività "peritale" e le regole stesse cambiano in funzione della disponibilità di nuove soluzioni tecnologiche, dei mutamenti culturali, del livello di alfabetizzazione informatica degli utenti.

Data questa natura mutevole, più opportuno sarebbe impiegare una strategia che vada oltre le esigenze della progettazione e costituisca un vero e proprio monitoraggio dell'interazione che gli utenti hanno con un sito Internet.

Il presente manuale costituisce una descrizione esaustiva di uno strumento specificatamente realizzato per implementare questo monitoraggio fornendo indicazioni rapide e continue sulla qualità dell'interazione tra utenti e sito. Più in particolare, Usability Evaluation (Us.E.) è un questionario multidimensionale

che consente di effettuare una diagnosi rapida dell'usabilità, individuando eventuali criticità per le quali possa essere necessario approfondire l'analisi (per esempio, attraverso specifici test di usabilità) per poi intervenire attraverso il re-design¹. Data la rapidità con cui consente di ottenere informazioni rilevanti, Us.E. è uno strumento adatto a essere impiegato non solo per valutazioni ad hoc, ma anche -e soprattutto- in maniera continuativa: un vero e proprio osservatorio sull'usabilità.

Il manuale è suddiviso in due parti: nella prima (capitoli 1 e 2) verrà introdotto il tema dell'usabilità dei siti Internet e della sua misura, soprattutto attraverso l'uso di questionari. Questa parte è stata scritta a beneficio di chi, seppur coinvolto nella progettazione e realizzazione di siti Internet, non possiede conoscenze specifiche nell'ambito dell'ergonomia cognitiva. La seconda parte del manuale è, invece, quella più operativa (capitoli 3, 4, 5) e descrive specificatamente il questionario, le sue proprietà psicometriche e le procedure di somministrazione e scoring. Anche in questi capitoli si è tentato, per quanto possibile, di trovare il giusto equilibrio tra rigore scientifico e semplificazione per i non addetti ai lavori. Alla fine del manuale il lettore troverà due utili appendici: la prima offre suggerimenti per la stesura di un rapporto di valutazione dell'usabilità, mentre la seconda fornisce una traduzione in inglese degli item da impiegare per la valutazione di siti con utenza internazionale.

¹ Per queste sue caratteristiche, Us.E. 2.0 è stato inserito dal Gruppo di Lavoro Usabilità (GLU) del Dipartimento della Funzione Pubblica nel Protocollo eGLU1.0 (e sue versioni successive) allo scopo di individuare le criticità di navigazione e di interazione dei cittadini con i siti delle Pubbliche Amministrazioni.

Avvertenza

Us.E è uno strumento nato per scopi scientifici e completamente gratuito. Non esiste una versione a pagamento, ma è possibile utilizzare questo strumento anche per scopi commerciali. Anche il presente manuale, che descrive le procedure di somministrazione e scoring della versione 2.0 del questionario, è distribuito gratuitamente in formato PDF. L'unico limite imposto agli utilizzatori è impiegare il questionario nella sua versione completa, senza aggiungere o eliminare item, usando le norme di taratura periodicamente aggiornate, riportando il tipo di licenza (CC BY-ND) e citando il presente volume in eventuali pubblicazioni e report, siano essi in forma cartacea o digitale².

Tutti coloro che, a vario titolo, sono coinvolti nella progettazione/ valutazione di siti Internet possono impiegare questo strumento e non occorre un titolo di studio specifico o l'iscrizione a un ordine professionale. Teoricamente, chiunque voglia valutare l'usabilità soggettivamente esperita di un sito Internet può impiegare Us.E. nella sua forma carta e matita o digitale. In pratica, però, è opportuno che i valutatori siano competenti nell'uso dello strumento e nell'interpretazione dei punteggi.

Anche se il sito di supporto al questionario (vedi capitolo 3) è dotato di una serie di vincoli che riducono la possibilità di commettere errori, non li annullano. Per esempio, sebbene lo scoring sia automatico e il grafico venga mostrato solo quando si raggiunge la numerosità minima consigliata, resta comunque la possibilità che il valutatore abbia scelto la norma di taratura sbagliata, oppure che abbia usato il link al questionario per

² Il corretto modo di citare la fonte è Di Nocera, F. (2013). *Usability Evaluation 2.0: Una descrizione (s)oggettiva dell'usabilità*. Roma: Ergoproject.

raccogliere dati su utenti non rappresentativi.

A partire dal 2012 esiste una procedura di certificazione come valutatore Us.E. subordinata a un esame teorico-pratico³. Lo scopo della certificazione non è, ovviamente, quello di creare una figura professionale, quanto piuttosto censire le persone in grado di impiegare in maniera opportuna lo strumento, fornire loro aggiornamenti specifici e segnalare l'esistenza di questa expertise a organizzazioni eventualmente interessate a impiegare Us.E. nel processo di valutazione dell'usabilità dei loro siti.

³ Per ulteriori informazioni sull'esame di certificazione è possibile consultare il sito <http://www.use2.it>.

1. Che cos'è l'usabilità

Quando si impiega il termine "usabilità"⁴ è difficile evitare di ripiegare su espressioni quali "facilità d'uso", "semplicità" o "intuitività". Tuttavia, il concetto non riesce a essere catturato in maniera esclusiva da nessuno di questi termini e, anche se spesso ci si accontenta di definire l'usabilità come facilità di apprendimento e di impiego di un artefatto, i modi in cui è possibile descrivere e misurare la qualità dell'interazione utente-tecnologia sono molteplici.

Anche allo scopo di mettere ordine nell'abbondanza di definizioni disponibili in letteratura, la International Organization for Standardisation ha introdotto nel 1998 la norma ISO 9241⁵ in cui l'usabilità (specificatamente nella parte 11 della norma) è stata descritta come il grado in cui un artefatto può essere usato da particolari utenti per raggiungere determinati obiettivi con efficacia, efficienza e soddisfazione in uno specifico contesto d'uso.

Tipicamente, l'efficacia viene fatta coincidere con il raggiungimento degli obiettivi, l'efficienza con il tempo

⁴ L'introduzione del termine è tipicamente fatta risalire a Bennet (1979) che lo impiegò per identificare ciò che rende un dispositivo conveniente e pratico da usare da parte di un utente. Oggi esistono diversi testi e una gran quantità di articoli che trattano il tema dell'usabilità. La scelta di un testo di approfondimento dipende dal livello di conoscenza del lettore e dalle specifiche esigenze informative o formative. Un testo relativamente recente e che rappresenta un discreto compromesso tra chiarezza dell'esposizione e livello di approfondimento è "Designing with the Mind in Mind" (Johnson, 2010).

⁵ La norma si chiamava inizialmente "Ergonomic requirements for office work with visual display terminals", ma è stata recentemente aggiornata con il titolo "Ergonomics of Human-System Interaction".

necessario per eseguire un compito e la soddisfazione con le valutazioni soggettive degli utenti. Si tratta, ovviamente, di semplificazioni. Tuttavia, sebbene questa definizione "standard" non rappresenti che un compromesso tra diverse istanze teoriche (non fornisce chiare indicazioni operative), la norma evidenzia un aspetto molto importante e spesso eluso da diversi approcci teorici e procedurali: l'usabilità non è una caratteristica del prodotto in sé, ma dipende dalle caratteristiche dell'utente che lo utilizza, dall'obiettivo che l'utente intende raggiungere e dal contesto nel quale il prodotto viene usato. Implicitamente, la norma sottolinea dunque come l'usabilità non possa essere ricondotta alla presenza/assenza di caratteristiche, ma debba essere valutata tenendo in considerazione l'esperienza soggettiva dell'utente che, con le sue abilità e i suoi limiti, intende raggiungere un obiettivo mediante l'uso di una tecnologia e desidera farlo impiegando la minor quantità di risorse possibile (soprattutto non dovendone paradossalmente investire di più a causa della tecnologia che impiega). La progettazione, pertanto, non può prescindere dalla conoscenza delle esigenze, dei limiti e delle potenzialità degli utenti, nonché dall'attenta analisi dei compiti che si dovranno svolgere impiegando il dispositivo in un determinato contesto.

1.1. Il concetto di interfaccia

Non è difficile comprendere perché sia desiderabile progettare dispositivi che facilitino l'utente e rendano l'esperienza d'uso piacevole. Più complicato è comprendere perché sia importante parlarne. E' forse necessario spiegare a un calzolaio che le scarpe devono esser comode e durature? Il "problema"

dell'usabilità, in realtà, diviene tale solo successivamente la comparsa delle tecnologie informatiche ed è in quel contesto che da nozione implicita, l'usabilità diventa oggetto di interesse dei ricercatori e dei professionisti. Il problema va dunque inquadrato storicamente, in particolare per quel che concerne l'evoluzione delle interfacce. Naturalmente non è questa la sede adeguata per affrontare nel dettaglio la storia dell'interazione uomo-computer, cionondimeno occorre mostrare brevemente il percorso che ha determinato la necessità di una progettazione orientata all'utente, nonché il ricorso al costrutto di usabilità.

Con il termine interfaccia indichiamo ciò che, fraponendosi tra l'utente e la macchina, ne rende possibile l'uso. Noi, difatti, non facciamo funzionare una lavatrice, un tablet o un computer direttamente; in realtà, forniamo alla macchina un insieme di comandi attraverso dispositivi di input per modificare il suo funzionamento e ottenere informazioni sul suo stato attraverso dispositivi di output. L'importanza delle interfacce è ben rappresentata dall'osservazione che per l'utente l'interfaccia è il prodotto (Raskin, 2000). L'interfaccia è dunque il "traduttore" che consente la comunicazione tra individui e sistemi tecnologici; ma che linguaggio comprende (e in che linguaggio si esprime) il traduttore? Nel migliore dei mondi possibili, all'interfaccia si dovrebbe poter semplicemente parlare, così come facciamo con una persona. Recentemente il riconoscimento vocale è stato implementato con successo anche su smartphones e tablet, ma siamo ancora distanti da scenari fantascientifici in cui sia realmente possibile una comunicazione affidabile con un sistema tecnologico per mezzo del linguaggio naturale. Nei fatti, la comunicazione avviene attraverso dispositivi di input come tastiere e puntatori che manipolano simboli rappresentati

su un display (non fanno eccezione i sistemi touch in cui il dito rappresenta il puntatore e l'interazione avviene direttamente sullo schermo). Queste interfacce grafiche (Graphic User Interfaces o, più semplicemente GUI) facilitano enormemente l'uso di un sistema e per comprenderne il valore occorre tornare al tempo in cui la maggior parte dei computer ne erano privi.

Nelle prime macchine che siano mai state definite "computer" (ma che erano più simili a ingombranti calcolatrici che ai computer come li conosciamo oggi) l'interfaccia era molto scarna e, per fornire le istruzioni in un linguaggio comprensibile alla macchina, i loro utilizzatori impiegavano schede opportunamente perforate. Il livello di conoscenza del funzionamento del sistema da parte di questi pochi utenti esperti (spesso si trattava degli stessi progettisti) era tale da non rendere necessario alcun intervento mirato a facilitare l'interazione. Solo successivamente la maggior parte dei sistemi è stata dotata di interfacce a riga di comando (Command Line Interfaces o CLI) nelle quali un trattino intermittente sullo schermo (prompt) restava "in attesa" di un comando fornito dall'utente attraverso una tastiera. Con una CLI l'utente è lasciato a se stesso e alla sua conoscenza di una sintassi per comunicare con la macchina. La sintassi è alla portata di tutti, ma occorre impararla. In una tipica GUI, invece, siamo invitati a muoverci tra icone e menu che diventano molto rapidamente familiari e, per buona parte degli utenti, non sarà mai necessario ricorrere a un manuale di istruzioni. Con l'evoluzione delle interfacce, dunque, si riducono i livelli di "traduzione" a carico dell'utente e si riduce la conoscenza specialistica necessaria per usare la tecnologia informatica. Tuttavia, da quando lo sviluppo tecnologico ha consentito ai computer di fare la loro comparsa in contesti non specialistici,

la conoscenza dell'utente finale è diventata sempre più distante da quella del progettista. Ciò comporta che, se nel team di sviluppo non è prevista una figura professionale che abbia la competenza per tener conto di questa distanza, il prodotto sarà difficilmente utilizzabile o, addirittura, inutilizzabile. Dunque, per rendere efficace, efficiente e soddisfacente l'interazione con un artefatto, diventa di fondamentale importanza impiegare una logica di progettazione orientata all'utente: un approccio iterativo in cui il processo di valutazione dell'interazione con un artefatto fornisce importanti indicazioni per il suo continuo miglioramento fino al suo rilascio.

1.2. Usabilità dei siti Internet

Lo sviluppo di Internet ha senza dubbio rivoluzionato le modalità con le quali gli individui si rapportano alla realtà e numerose sono le operazioni che ormai le persone eseguono quasi esclusivamente impiegando questo mezzo: controllare il conto in banca, effettuare acquisti, organizzare un viaggio, prenotare posti a teatro o al cinema, come pure mantenere relazioni con persone distanti e promuoversi professionalmente. Inoltre, se in principio Internet era semplicemente un mezzo per scambiare informazioni, oggi – attraverso social networking, blogging e micro-blogging – offre la possibilità di esprimere (e orientare) giudizi e opinioni anche su importanti questioni di natura sociale e politica. Basti pensare all'operazione "Obama's Online Operation" (Triple O) del 2008 che ha consentito la raccolta di 600 milioni di dollari di fondi a sostegno della campagna da parte degli elettori, oppure alla campagna elettorale 2013 del Movimento Cinque Stelle in Italia che è stata condotta quasi

esclusivamente su web. La progettazione di siti Internet “usabili” è quindi diventata non solo un fattore di miglioramento generale delle interfacce ma anche un elemento di importanza strategica per la competizione professionale, aziendale, politica.

A questo punto, dato il focus sull’usabilità dei siti Internet, è opportuno rispondere brevemente a una delle domande che spesso viene posta quando si affronta questo argomento, vale a dire se sia necessario trattare l’usabilità dei siti diversamente dall’usabilità del software. Anni fa la domanda era giustificata dal fatto che il principale oggetto di studio era proprio l’interazione con il software (sistemi operativi inclusi). Erano anni in cui la comunità di ricercatori che si riconosceva nell’etichetta “Human-Computer Interaction”⁶ era già affermata, ma il web non era ancora pervasivo. Più recentemente la domanda invece è giustificata dalla diffusione di “App” che sono tecnicamente degli applicativi software, ma che in quanto a numerosità e varietà pongono problemi analoghi all’interazione con un sito.

Per provare a rispondere a questa domanda occorre quindi in primo luogo fare una precisazione: tecnicamente possiamo valutare l’usabilità di qualunque cosa e non c’è ragione di credere che la cornice teorica per valutare l’usabilità di una sedia debba essere differente da quella impiegata per valutare l’usabilità di uno smartphone o di un pannello di controllo in una centrale nucleare. Ciò che fa la differenza, nelle diverse circostanze, è piuttosto la salienza di alcuni aspetti dell’interazione rispetto ad

⁶ L’espressione è diventata popolare con la pubblicazione libro “The Psychology of Human Computer Interaction” (Card, Moran e Newell, 1983).

altri. In alcuni casi potrebbe essere rilevante l'attività motoria (si pensi alle "gestures" impiegate su smartphones e tablet), in altri la leggibilità del testo o la comprensibilità delle voci di menu. Aspetti differenti, dunque, che diventeranno rilevanti nel momento in cui sarà opportuno individuare i cambiamenti necessari per migliorare l'interazione, ma che risulterebbero fin troppo specifici rispetto allo scopo di ottenere dall'utente una stima della qualità della sua esperienza di interazione. Dunque, risulterebbe di modesta utilità parlare dell'usabilità dei siti Internet quasi fosse materia diversa dall'usabilità del software. La differenza tra siti Internet e software risiede piuttosto nelle conseguenze della qualità dell'interazione. Un pacchetto software (per esempio, un software per condurre analisi statistiche) costituisce uno strumento con il quale occorre necessariamente familiarizzare, in alcuni casi perché lo impone l'organizzazione presso la quale l'utente lavora, in altri casi perché l'utente stesso ha sostenuto i costi di acquisto, in altri ancora perché la disponibilità di alternative è ridotta o assente. Tutto ciò rende poco probabile l'abbandono e la sostituzione della tecnologia. Per un sito internet, invece, un'esperienza di interazione insoddisfacente può determinare rapidamente l'abbandono perché c'è abbondanza di siti alternativi per qualsivoglia esigenza.

L'esperienza dell'utente è quindi un fattore critico nella progettazione di siti Internet e non può essere limitata a definire le caratteristiche strutturali o estetiche dell'interfaccia. Occorre piuttosto creare un ponte tra la domanda dell'utente e la risposta fornita dal sito, perché le persone non visitano un sito per il gusto di farlo, ma per soddisfare le loro esigenze (informative, d'acquisto, ecc.). A tal proposito, Travis (2006) ha impiegato la

metafora delle “red routes” londinesi per indicare la necessità di individuare percorsi preferenziali in linea con la mission del sito e gli obiettivi di navigazione dell’utente. Esattamente come i percorsi progettati per facilitare lo scorrimento del traffico, chiaramente riconoscibili e in cui è vietato sostare, un sito dovrebbe essere dotato di percorsi che consentano di partire dalle esigenze degli utenti o dagli obiettivi dell’organizzazione e di raggiungere un punto di arrivo rappresentato dalla soddisfazione di quelle esigenze. Per esempio, uno degli obiettivi dei clienti di un sito e-commerce potrebbe plausibilmente essere quello di assicurarsi il miglior rapporto qualità-prezzo. Una strategia per soddisfare questa necessità potrebbe essere quella di introdurre un’area con collegamenti a recensioni del prodotto, commenti di altri clienti e i prezzi proposti da siti concorrenti. Una “red route” rappresenta dunque un’attività completa (non una serie di compiti distinti) e progettare tenendo conto di questi percorsi può determinare il successo di un sito. In altre parole, investire in usabilità comporta un immediato ritorno economico che si esprime non solo in termini di guadagno, ma anche di risparmio. Evitare che gli utenti possano non accettare -e dunque decidere di non utilizzare- un servizio online perché non si è investito a sufficienza su una progettazione orientata all’utente, rappresenta un risparmio notevole. Questo aspetto non riguarda solo i siti commerciali dove l’usabilità è già riconosciuta come un beneficio tangibile, ma anche quelli degli Enti, delle Pubbliche Amministrazioni, delle Università.

2. Misurare l’usabilità

Esistono almeno due buoni motivi per misurare l’usabilità dei

siti Internet. Il primo dovrebbe ormai essere chiaro: l'usabilità è un elemento critico per definire il successo o il fallimento di un servizio on-line e non rappresenta una qualità accessoria di cui occuparsi a valle della progettazione e della realizzazione del sito. Il secondo motivo è intimamente legato alla necessità di fornire un controllo di qualità laddove la comunicazione via Internet coinvolga pubbliche amministrazioni, enti pubblici e istituti di formazione come scuole e università. In questi casi, difatti, la comunicazione on-line di iniziative e atti costituisce un adempimento ineludibile ma che, sfortunatamente, può essere assolto anche solo rendendo disponibili informazioni e documenti senza curarsi degli aspetti di interazione⁷.

Il questionario Usability Evaluation 2.0 che viene proposto in questo manuale consente di ottenere indicazioni rapide (e all'occorrenza aggiornate nel tempo) sull'usabilità stimata dagli utenti stessi. Naturalmente, non si tratta dell'unico metodo a disposizione del valutatore e neppure l'unico questionario disponibile. Una qualunque analisi, anche superficiale, della letteratura sul tema dell'usabilità è in grado di mostrare quanto sia variegata l'offerta di metodi e strumenti. Da quando è stata introdotta la nozione di usabilità, difatti, la comunità di ricercatori e professionisti che se ne occupa si è enormemente sviluppata. Ciò ha comportato un conseguente proliferare di letteratura specialistica e divulgativa, associazioni scientifiche

⁷ Il principio della trasparenza, affermato attraverso il decreto legislativo 33/2013, fa riferimento al concetto di "accessibilità" delle informazioni e l'obiettivo della norma è unicamente quello di favorire un controllo diffuso da parte del cittadino sull'operato delle istituzioni e sull'utilizzo delle risorse pubbliche. In pratica, la presenza online è stata imposta senza rendere altrettanto obbligatoria una procedura per controllarne la qualità.

e professionali, corsi di formazione e certificazioni, ma anche di modelli e metodi che costituiscono il corpus di conoscenze necessario per chiunque voglia intraprendere una carriera in questo settore. Senza lasciarci ingannare dalle etichette nelle quali i diversi professionisti si riconoscono (Interaction Designer, User Experience Specialist, Usability Analyst, Information Architect, ecc.), i metodi necessari per indagare e migliorare l'interazione tra individui e tecnologie non differiscono in funzione della qualifica che il singolo decide di associare al proprio nome sul biglietto da visita.

D'altra parte, il tema della misura dell'usabilità è piuttosto complesso. Si tratta di un argomento vasto e pieno di insidie, sia perché la posizione teorica, la tipologia di prodotto e la fase di sviluppo determinano la scelta della tecnica da impiegare, sia perché esiste un certa sovrapposizione (e confusione) di ruoli tra la comunità scientifica e quella professionale che non rende facile al neofita distinguere tra metodi "quick & dirty"⁸ e metodi/strumenti validi e affidabili.

Le sezioni successive prenderanno brevemente in considerazione i due grandi approcci⁹ per misurare l'usabilità: quelli che non coinvolgono gli utenti e quelli che li coinvolgono. Si tratta di una visione personale e critica mirata soprattutto a spiegare, in ultima analisi, perché venga suggerito l'uso

⁸ Con l'espressione "quick & dirty" si indica qualunque soluzione ritenuta accettabile per gli scopi di chi la implementa. In questo specifico caso, è riferita a questionari realizzati ad hoc e senza far ricorso a strategie di analisi e di validazione articolate e scientificamente fondate.

⁹ Per una descrizione delle tecniche citate in questo paragrafo si rimanda a Di Nocera (2004).

dei questionari nel processo di valutazione (e limitatamente alla valutazione¹⁰). Inoltre, non è superfluo ricordare che le considerazioni che seguono fanno riferimento all'uso di determinate strategie di ricerca/valutazione nel contesto specifico dei siti Internet.

2.1 Stima dell'usabilità senza il coinvolgimento degli utenti

Tra gli approcci che non prevedono il coinvolgimento di utenti, possiamo distinguere quelle tecniche basate sulla sequenza logica delle operazioni che un utente ideale dovrebbe eseguire interagendo con il sistema (metodi formali basati su una task analysis) e quelle basate su liste di funzionalità che, se presenti, renderebbero usabile un sistema (valutazioni euristiche condotte da giudici esperti). In entrambi i casi, lo "strumento di misura" è l'esperto o il team di esperti che guida la valutazione.

È evidente che questi approcci non indagano l'interazione tra utenti e tecnologie, l'oggetto di indagine è piuttosto un modello di utente o un modello di sistema. Esistono buone ragioni per adottare queste tecniche in talune circostanze, si tratta in fin dei conti di metodi a basso costo e in grado di orientare il design nelle primissime fasi di sviluppo. Tuttavia, questi approcci non

¹⁰ La presente trattazione riguarda il tema della misura, non quello dell'ideazione e della progettazione. Alcune tecniche sono indispensabili per creare interfacce usabili, ma non forniscono informazioni utili alla misurazione del costruito. Per esempio, il Card Sorting è tipicamente impiegato per creare l'architettura di un sito che sia compatibile con la rappresentazione dei contenuti che gli utenti mostrano di avere. L'uso della tecnica, pur promettendo una maggiore usabilità del sito, non ne costituisce una misura

prendono in considerazione né il livello di expertise degli utenti reali che utilizzeranno il sistema, né i loro obiettivi.

Una strategia simile alla valutazione da parte di esperti consiste nell'impiego di strumenti automatici in grado di rilevare la presenza/assenza di caratteristiche che renderebbero più o meno usabile un sito. Per quanto possa essere attraente la possibilità di affidarsi a un servizio automatico per ottenere rapidamente un "punteggio" di usabilità, questa pratica dovrebbe essere fortemente scoraggiata. Nessun algoritmo, difatti, è in grado di risalire alle problematiche di interazione che gli utenti di uno specifico sito potrebbero rilevare. Questi servizi possono al massimo rivelarsi di una qualche utilità per identificare cattive pratiche ed errori grossolani (per esempio, l'aver usato un font rosso su uno sfondo blu) che ai giorni nostri dovrebbero essere piuttosto improbabili¹¹.

2.2 Stima dell'usabilità con coinvolgimento degli utenti

Se rispettiamo la definizione di usabilità fornita nel primo capitolo, per ottenere indicazioni sulla qualità dell'interazione ci sarà ovviamente bisogno di avere a disposizione tutti gli attori coinvolti. Se l'obiettivo è quello di valutare la qualità dell'interazione con un sito Internet, occorrerà quindi avere

¹¹ Strumenti simili vengono comunemente utilizzati per stimare se il codice è in grado di assicurare l'accessibilità di un sito, vale a dire la possibilità per utenti permanentemente o temporaneamente disabili (per esempio non vedenti o ipovedenti) di accedere in maniera alternativa ai contenuti. L'uso di strumenti di valutazione automatica non è esente da critiche anche in questi casi.

accesso al sito in esame e ai suoi utenti reali. Se il sito ancora non esiste si potranno utilmente impiegare prototipi (cartacei o digitali) e utenti potenziali. In ogni caso, ciò non può prescindere dal fatto che l'interazione debba avvenire e che alcune variabili chiave vengano misurate. Anche nel caso degli approcci che prevedono il coinvolgimento degli utenti è possibile distinguere, a grandi linee, alcune famiglie di tecniche: quelle in cui i compiti di navigazione del sito vengono indicati dal valutatore e quelle in cui la valutazione è basata sull'analisi dell'attività spontanea dell'utente. In pratica, le prime sono (oppure somigliano a) esperimenti di laboratorio, mentre le seconde corrispondono all'osservazione naturalistica.

Anche se collettivamente questi approcci vengono definiti "test di usabilità", non hanno una specificità per lo studio di questo costrutto, ma rappresentano adattamenti di strategie ampiamente impiegate in Psicologia per studiare il comportamento umano. Le variabili tipicamente indagate sono il tempo impiegato per portare a termine i compiti, come pure il numero e la tipologia di errori commessi. Si tratta di allestire piccoli esperimenti in cui un campione di utenti debba portare a compimento alcune operazioni considerate critiche (per esempio, il ritrovamento di una specifica informazione o la procedura di acquisto di un prodotto). Durante queste sessioni l'utente potrà eseguire compiti precedentemente definiti dallo sperimentatore, oppure essere lasciato libero di interagire come meglio crede mentre viene direttamente osservato o video-registrato. In quest'ultimo caso, l'osservatore prenderà nota delle strategie impiegate dall'utente, degli errori commessi ed eventualmente porrà domande specifiche per comprendere i motivi delle scelte operate.

Realizzare un esperimento è un'attività complessa che necessita di competenze metodologiche e statistiche che spesso è implausibile richiedere al consulente che si occupa di usabilità. Per questo motivo, i test che vengono introdotti dai "guru" della UX sono confezionati per essere alla portata di chi non ha un'esperienza di ricerca sperimentale. Inoltre, sono test che vengono proposti per essere utilizzati impiegando pochi soggetti e senza far ricorso a strumenti statistici per sottoporre a verifica le ipotesi. Dunque, nonostante derivino dalla ricerca sperimentale, i test di usabilità molto spesso non ne mostrano il rigore. Discorso simile andrebbe fatto per i metodi "etnografici" che derivano dal contesto dell'analisi sociologica e antropologica ma che, avendo obiettivi e tempi differenti, non rappresentano che una pallida imitazione di quelle metodologie.

Naturalmente, queste osservazioni non sono mirate a scoraggiare l'attività professionale, ma a suggerire particolare attenzione nella scelta dei metodi da impiegare e nel verificare l'utilità delle informazioni che essi forniscono. Per esempio, dal momento che la navigazione di un sito Internet consente di tener traccia sul server dei comportamenti del visitatore (durata della sessione, percorso di esplorazione, ecc.), uno dei metodi lungamente utilizzato per studiare i comportamenti di navigazione è stato l'analisi dei logfile (o, più semplicemente, log). Tuttavia, questa analisi -per quanto fonte utilissima di informazioni- poco ci dice sulle ragioni del comportamento dell'utente. Si tratta di una tecnica meramente descrittiva del percorso seguito e, in condizioni non controllate (condizioni in cui non stiamo osservando l'utente in laboratorio), le informazioni desunte potrebbero addirittura essere fuorvianti. Per esempio, un utente potrebbe sostare su una pagina per

diverse ragioni: potrebbe trovare particolarmente interessante i contenuti, potrebbe trovare quei contenuti di difficile comprensione, oppure potrebbe semplicemente aver lasciato la pagina aperta durante una telefonata. Allo stesso modo, tornare frequentemente al menu principale potrebbe essere considerato rappresentativo di una difficoltà nell'interazione, mentre l'utente potrebbe essere semplicemente in compagnia di una persona alla quale sta mostrando il sito.

Gli approcci sopra riportati hanno un elemento critico in comune: dipendono dalla domanda specifica che il ricercatore o consulente si pone rispetto al sito in esame. Non rappresentano dunque misure dell'usabilità, ma strategie (sperimentali, di osservazione e di analisi) per misurare particolari variabili che si ritengono indicative dell'usabilità, facendo ricorso a tecniche e strumenti tipicamente impiegati nella ricerca psicologica (tempi di risposta, numero di errori, strategie di esplorazione visiva¹², ecc.). Questo modo di procedere è utilissimo per approfondire specifici aspetti dell'interazione utente-sito, ma del tutto sovradimensionato qualora l'esigenza fosse quella di ottenere una stima dell'usabilità di un sito Internet.

L'ultima strategia di misurazione che prenderemo in considerazione è quella oggetto del presente manuale. La prossima sezione è difatti dedicata ai questionari.

¹² Il tracciamento dei movimenti oculari (eye-tracking) sta diventando molto popolare come metodologia di indagine in questo settore. Tuttavia, la registrazione dei movimenti oculari non costituisce di per sé una misura di usabilità e per comprendere il significato funzionale dei pattern di esplorazione è fondamentale l'abilità del ricercatore o del consulente nell'analizzare i dati e interpretare i risultati opportunamente.

2.3 L'uso dei questionari nella valutazione dell'usabilità

I questionari hanno l'indubbio merito di essere uno dei più veloci ed economici strumenti che si possano impiegare per ottenere informazioni direttamente dall'utente. Tuttavia, un questionario potrebbe sembrare uno strumento troppo debole per quantificare l'usabilità. Per molte persone, i questionari rappresentano un modo per raccogliere opinioni e non costituiscono strumenti di misura veri e propri. A beneficio dei non addetti ai lavori, occorre dunque fare una distinzione molto importante tra semplici liste di domande e strumenti psicometrici. Per essere considerato uno strumento, un questionario deve possedere alcune proprietà fondamentali. In primo luogo deve essere valido, vale a dire che ogni domanda o affermazione deve "catturare" il costrutto teorico in esame (o un aspetto di quel costrutto nel caso di questionari multi-dimensionali). Un'altra proprietà imprescindibile per un questionario che possa considerarsi uno strumento di misura è la sua attendibilità, vale a dire la capacità di fornire misurazioni coerenti.

Inoltre, occorre ricordare che, a differenza di altri metodi di misurazione, un questionario fornisce informazioni rispetto a come l'utente percepisce una determinata situazione. È dunque opportuno limitare l'uso delle misure soggettive solo a quegli aspetti che non possono essere indagati in altro modo, oppure laddove la percezione dell'utente è esattamente l'oggetto di studio (come in questo caso).

La realizzazione di un questionario non consiste dunque nel mettere insieme un elenco di domande o affermazioni. Creare un questionario è molto laborioso, comporta fasi successive di

testing degli item scelti e l'uso di specifiche tecniche di analisi per assicurarsi che possieda le proprietà di attendibilità e validità che ci aspettiamo da uno strumento di misura.

Sfortunatamente, non tutti i questionari proposti come misure di usabilità rispettano queste caratteristiche, anche se la loro diffusione potrebbe lasciare intendere il contrario. Peraltro, alcuni questionari sono nati con altri scopi, ma hanno ottenuto la fama di strumenti utilizzabili anche oltre la loro destinazione originale. Non è raro, quindi, leggere report di valutazione, articoli e libri che li citano o li impiegano come misura dell'usabilità dei siti Internet, anche se la loro realizzazione non prevedeva questo impiego. Spesso l'adattamento al web è stato operato aggiungendo, eliminando o trasformando le domande e le affermazioni che compongono un questionario con una disinvoltura che lascia intendere una sottovalutazione della stima soggettiva del costrutto, oppure un'ignoranza dei presupposti di validità e attendibilità necessari. Per esempio, il Questionnaire for User Interaction Satisfaction (QUIS: Chin, Diehl e Norman, 1988) - nato per valutare l'interazione con altri tipi di sistemi e dichiaratamente una misura di soddisfazione più che di usabilità- è uno degli strumenti più saccheggianti. Altri strumenti, come il Website Analysis and MeasureMent Inventory (WAMMI: Kirakowsky et al., 1998), rappresentano invece estensioni o adattamenti al web di questionari precedentemente sviluppati per la valutazione dell'interazione con sistemi software. Discorso a parte merita la System Usability Scale (SUS: Brooke, 1986), che viene impiegata come "misura" globale dell'usabilità indipendentemente dal sistema oggetto di valutazione. L'uso frequente di uno strumento lo rende uno "standard" de facto, indipendentemente dalla sua bontà. Non

è sorprendente, quindi, che questa scala, composta di 10 item poco indicativi dell'esperienza dell'utente di un sito Internet e addirittura non applicabili alla realtà oggetto di studio¹³, sia stata citata in migliaia di pubblicazioni e venga ininterrottamente utilizzata dal 1986 per valutare l'usabilità di sistemi hardware, software e siti Internet.

Il prossimo capitolo descriverà il questionario Usability Evaluation, il cui sviluppo è stato curato impiegando gli stessi criteri e lo stesso rigore che ha caratterizzato lo sviluppo di altri test di comprovata validità e affidabilità. Ciò attiene gli aspetti metodologici. Da un punto di vista teorico, invece, la principale differenza tra Us.E. e altri questionari per la valutazione dell'usabilità consiste nel partire dal presupposto che il costruito in esame debba essere rappresentato unicamente in termini di valutazioni soggettive dell'utente che ha interagito con un sistema. Us.E. dunque si propone come un metodo per diagnosticare l'usabilità dei siti Internet, lasciando alle altre strategie -brevemente passate in rassegna nelle sezioni precedenti- il compito di fornire informazioni necessarie per la progettazione e la ri-progettazione.

¹³ L'item n° 4, per esempio, recita "Penso di aver bisogno del supporto di un tecnico per utilizzare questo sistema" (*I think that I would need the support of a technical person to be able to use this system*). Lascio al lettore stabilire quanto spesso abbia pensato di aver bisogno del supporto di un tecnico per usare un sito Internet.

3. Il Progetto Us.E.

Usability Evaluation (UsE.) è un questionario multi-dimensionale che permette di stimare l'usabilità dei siti Internet, intesa come qualità dell'interazione esperita da un utente che visita un sito per raggiungere un obiettivo (per esempio, ricerca di informazioni, acquisto di un prodotto, compilazione di un form). Us.E. consente di effettuare una diagnosi rapida dello "stato di salute" di un sito, individuando le eventuali aree problematiche rispetto le quali sarà opportuno approfondire l'analisi. La valutazione viene effettuata sulla base di tre dimensioni, definite "Maneggevolezza", "Soddisfazione" e "Attrattiva", che verranno presentate nel dettaglio più avanti.

3.1 Timeline

Lo sviluppo del questionario è iniziato nel 1997, ma lo strumento è stato presentato per la prima volta solo due anni dopo, in occasione della 1^a Giornata italiana su Human-Computer Interaction (Di Nocera, Ferlazzo, Renzi, 1999). Il primo studio aveva impiegato una versione del questionario che conteneva numerose affermazioni precedentemente ottenute attraverso interviste a utenti di diversi siti Internet che si erano espressi sulla loro esperienza d'uso.

In particolare, nella sua versione iniziale (Us.E. 1.0), il questionario era composto di 70 item e le analisi condotte avevano suggerito l'esistenza di una struttura a 4 fattori: "Maneggevolezza", "Soddisfazione", "Attrattiva" e "Prevedibilità" (Di Nocera, Ferlazzo, Renzi, 2003). Successive analisi hanno

suggerito una soluzione più parsimoniosa a tre fattori, dove gli item facenti riferimento alla dimensione della Prevedibilità confluivano nella dimensione della Maneggevolezza. Questa versione è stata quindi distribuita con un primo set di norme di taratura¹⁴ e impiegata per scopi di ricerca e professionali.

Un elemento critico di questa versione era l'elevato numero di item, che rendeva Us.E. poco agevole per la somministrazione on-line (Boscarol, 2003). A questo scopo è stata condotta una raccolta dati che ha coinvolto un elevato numero di siti (valutati da migliaia di utenti) al fine di individuare una struttura sufficientemente stabile e selezionare gli item¹⁵ che maggiormente contribuivano a determinarla.

3.2 Dimensioni

Le dimensioni di Us.E. sono state definite secondo una logica bottom-up, vale a dire che sono state individuate a partire dai dati e non in funzione di un modello teorico di riferimento. Contrariamente a ciò che accade per molti questionari, assemblati sulla base del buon senso o dell'esperienza e adottati per specifiche necessità, i risultati delle analisi condotte per la

¹⁴ Le norme di taratura sono punteggi di riferimento che permettono di confrontare i risultati ottenuti sul singolo sito con quelli ottenuti valutando altri siti appartenenti alla stessa categoria.

¹⁵ A partire dalla sua introduzione, alcune versioni intermedie di Us.E. sono state sviluppate e impiegate per giungere a una forma ritenuta soddisfacente. Tra queste, la versione 1.1 era composta di 24 item ed è stata per lungo tempo utilizzata prima di rilasciare la versione corrente e "stabile" del questionario. Il criterio di numerazione delle versioni segue quello impiegato per le release software (numeri dispari per le versioni di sviluppo, numeri pari per quelle stabili).

creazione e la validazione di Us.E. sono documentati (Di Nocera, Ferlazzo, Renzi, 2003; Terenzi, Di Nocera, Ferlazzo, 2006; Di Nocera, Terenzi, Ferlazzo, 2007; Di Nocera et al., 2009) e il prossimo capitolo descriverà le caratteristiche psicometriche dello strumento.

Di seguito vengono brevemente presentate le tre dimensioni lungo le quali viene fornita la stima di usabilità e i relativi item. Nella sua versione più recente Us.E. 2.0 è composto di 19 affermazioni e agli utenti viene richiesto di esprimere il proprio giudizio su ciascuna di esse utilizzando una scala Likert a 5 punti (da "assolutamente falso" ad "assolutamente vero").

3.2.1 Scala M (Maneggevolezza)

Questa dimensione fa riferimento all'interazione con le proprietà strutturali del sito, quindi con la sua architettura o organizzazione e con i menu di navigazione. In generale, esprime la facilità con cui l'utente è in grado di raggiungere l'obiettivo che si è prefissato visitando un determinato sito. Gli item che definiscono questa dimensione sono undici, di cui quattro esprimono la dimensione negativamente (assenza di maneggevolezza) e pertanto andranno ricodificati¹⁶ in fase di

¹⁶ Alcuni item definiscono negativamente la dimensione in esame. Per esempio, "Ho provato una sensazione di smarrimento esplorando questo sito" è chiaramente indicativo della mancanza di Maneggevolezza. In questi casi occorrerà invertire la scala di risposta: vale a dire, se l'utente ha risposto 1, il punteggio dovrà essere trasformato in 5. In tal modo, le risposte agli item appartenenti alla stessa scala avranno la stessa direzione e potranno essere sommate.

scoring (vedi capitolo 5).

Punteggi bassi in questa scala suggeriscono la necessità di affrontare cambiamenti strutturali quali la riorganizzazione della gerarchia dei contenuti, la posizione degli elementi all'interno delle pagine o la modifica del menu di navigazione.

1.	Esplorando questo sito capisco ogni momento dove mi trovo.
3.	Fin dall'inizio è chiaro quali sono i contenuti di questo sito.
5.	Ho provato una sensazione di smarrimento esplorando questo sito.
7.	Posso raggiungere l'indice generale degli argomenti con facilità.
9.	Questo sito è difficile da consultare.
11.	Visitare questo sito è stato facile tanto quanto usare il software che uso di più.
12.	In questo sito posso trovare quello che mi interessa senza doverlo esplorare tutto.
14.	In questo sito sono stato sul punto di perdermi.
16.	Ho un controllo continuo sulle operazioni possibili in questo sito.
17.	Le informazioni presenti in questo sito sono comprensibili.
19.	Questo sito presenta lunghe liste di difficile consultazione.

Tabella 1. Item della scala "Maneggevolezza" con la loro numerazione originale.

3.2.2 Scala S (Soddisfazione)

Questa dimensione fa riferimento alla "utilità percepita"¹⁷ del sito, dunque è intimamente legata al raggiungimento degli obiettivi dell'utente o, detto in altri termini, alla soluzione del problema che ha richiesto l'uso di quello specifico sito. Gli item che contribuiscono a questa scala sono sei, di cui tre esprimono la dimensione negativamente (assenza di soddisfazione) e pertanto andranno ricodificati in fase di scoring.

Punteggi bassi in questa scala potrebbero indicare che il sito non è stato realizzato tenendo conto del suo bacino di utenza. Ciò può accadere per almeno due motivi: l'utenza non è quella attesa da chi ha realizzato il sito, oppure i contenuti non sono adeguati alle aspettative degli utenti.

2.	Questo sito ha deluso le mie aspettative.
4.	Questo sito è tanto bello quanto inutile.
8.	Questo sito è inutile con la parvenza di essere utile.
13.	I contenuti di questo sito sono aggiornati.
15.	Sono riuscito ad ottenere le informazioni/servizi che cercavo.
18.	Esplorare questo sito è stata una perdita di tempo.

Tabella 2. Item della scala "Soddisfazione" con la loro numerazione originale.

¹⁷ Il concetto di soddisfazione di cui ci occupiamo in questa sede potrebbe essere confuso con quello di "customer satisfaction" impiegato nel marketing. In realtà si tratta di due aspetti ben distinti. La customer satisfaction fa riferimento

3.2.3 Scala A (Attrattiva)

Questa dimensione fa riferimento all'apprezzamento di alcune caratteristiche estetiche (colore e composizione) impiegate dall'utente per ricavare informazioni indirette sulla cura che è stata impiegata per realizzare il sito. Queste informazioni sono in grado di veicolare informazioni sull'usabilità ancor prima di interagire con il sito (la cosiddetta "usabilità apparente": Kurosu e Kashimura, 1995). Gli item che contribuiscono a questa scala sono solo due, entrambi definiscono positivamente la dimensione.

Punteggi bassi potrebbero indicare la necessità di un restyling e di maggiore cura nei particolari.

6.	La scelta dei colori utilizzabili in questo sito è elegante.
10.	Gli elementi grafici impiegati sono accattivanti e curati nei particolari.

Tabella 3. Item della scala "Attrattiva" con la loro numerazione originale.

3.3 Anagrafica e opinioni degli utenti

Oltre agli item che compongono le scale, Us.E. 2.0 è accompagnato da una parte anagrafica per ottenere informazioni sui rispondenti. In particolare, vengono richiesti il genere, l'età, il livello di scolarità, la frequenza d'uso di Internet (quotidiana vs. occasionale) e la pressaggia conoscenza del sito

alla proporzione di utenti soddisfatti di un prodotto o di un servizio rispetto a specifici obiettivi aziendali, mentre il concetto di soddisfazione che trattiamo in tema di usabilità fa riferimento più all'utilità percepita dall'utente nella sua esperienza di navigazione con il sito.

(prima visita vs. visita successiva). Queste variabili sono utili per segmentare i risultati e per interpretare i punteggi ottenuti attraverso le tre scale.

Al termine del questionario, è presente uno spazio nel quale l'utente ha la possibilità di lasciare la propria opinione sul sito con il quale ha appena interagito. Molti utenti approfittano di questa opportunità e il materiale raccolto è di importanza strategica per meglio comprendere le criticità del sito in esame e arricchire il report di valutazione (vedi Appendice 1).

3.4 Licenza d'uso

Us.E. è rilasciato con licenza CC BY-ND, dove gli attributi BY e ND stanno rispettivamente a significare che 1) bisogna sempre indicare l'autore dell'opera e che 2) non sono consentite elaborazioni dell'opera creativa. Una licenza Creative Commons (CC) rappresenta una via di mezzo tra il modello del copyright © in cui tutti i diritti sono riservati e quello di pubblico dominio (PD) in cui nessun diritto è riservato. In questo caso alcuni diritti sono riservati. L'organizzazione Creative Commons (<http://www.creativecommons.org>) ha pubblicato diversi tipi di licenze che permettono ai creatori di scegliere e comunicare quali diritti riservarsi e a quali rinunciare.

Riepilogando, il questionario, la procedura di scoring e i valori normativi possono essere liberamente utilizzati a patto che 1) venga citata la fonte e 2) non vengano modificati gli elementi costitutivi del questionario, la procedura di scoring e i valori normativi per calcolare i punteggi standardizzati.

3.5 Controindicazioni

Come è stato riportato nel capitolo 2, nella pratica professionale alcuni questionari vengono impiegati con disinvoltura per valutare l'usabilità di sistemi per i quali non sono stati specificatamente realizzati. E' opportuno ribadire che Us.E. 2.0 è un questionario ideato per valutare l'usabilità dei siti Internet e non è adatto per altri tipi di sistemi. Il progetto Us.E., difatti, nasce indagando la relazione tra utenti e ipertesti, vale a dire insieme di pagine residenti su un server e messi in relazione tra loro per mezzo di collegamenti. Tecnicamente, tutti i siti Internet sono ipertesti, ma i social network e i sistemi di micro-blogging rappresentano un'evoluzione inaspettata e non coincidono pienamente con la tipologia di sito che il questionario intende valutare¹⁸. L'impiego di Us.E. per valutare cose diverse dai siti Internet "tradizionali" è quindi non solo sconsigliato, ma totalmente errato.

¹⁸ Se si impiegasse Us.E. per valutare un social network, l'oggetto della valutazione sarebbe diverso per ogni utente. I contenuti, infatti, sono personalizzati in funzione dei feed cui l'utente ha sottoscritto e delle regole di selezione dei contenuti che l'utente ha scelto di implementare.

4. Proprietà psicometriche

4.1 Campione

Nel corso degli anni sono stati analizzati molti siti e raccolti dati di numerose valutazioni impiegando diverse versioni del questionario¹⁹. In principio sono stati analizzati siti universitari e, successivamente, la ricerca è stata ampliata ad altre categorie, mostrando l'esistenza di differenze notevoli tra le valutazioni effettuate sulle diverse tipologie di siti. Chiaramente, le valutazioni ottenute attraverso Us.E. necessitavano di essere riferite alla giusta categoria di appartenenza di un sito. Tuttavia, categorizzare è un processo difficile, in quanto ogni categorizzazione è in una qualche misura arbitraria. Malgrado le difficoltà e la dose di arbitrio necessaria, organizzare i dati in funzione della tipologia di sito è stato un passo indispensabile al fine di fornire dei valori normativi rispetto ai quali confrontare i dati provenienti da una singola valutazione. Le norme di taratura consentono infatti di contestualizzare i punteggi ottenuti dal sito oggetto di valutazione rispetto la categoria di appartenenza. Questa procedura di contestualizzazione è fondamentale per l'interpretazione dei punteggi. Si prenda come esempio un valore pari a "8" nella scala Attrattiva (scala per la quale il punteggio massimo ottenibile è pari a 10). Questo punteggio potrebbe costituire un valore medio qualora si trattasse di un sito commerciale (dal momento che questa

¹⁹ Un totale di 27 siti (valutati da 2136 utenti) hanno contribuito al pool di dati che è stato impiegato per individuare la struttura fattoriale e fornire le prime norme di taratura della versione 2.0 del questionario.

tipologia di siti mira a essere più accattivante e i punteggi medi di riferimento potrebbero essere più prossimi a 10). Lo stesso punteggio potrebbe, invece, rappresentare una valutazione molto positiva se il sito in questione fosse quello di una Pubblica Amministrazione, perché questa tipologia di siti è generalmente molto spartana e viene valutata come poco accattivante da parte degli utenti.

È quindi indispensabile riferire il sito oggetto di valutazione a una categoria con la quale poter confrontare il risultato ottenuto o, in ogni caso, a quella più vicina dal punto di vista tematico.

Le categorie finora individuate²⁰ -e per le quali sono state rese disponibili le norme di taratura- non sono definitive, costituiscono solo un punto di partenza e potrebbero con il tempo essere modificate in funzione della disponibilità di dati e della individuazione di differenze significative tra sotto-tipologie di siti. Per esempio, è del tutto plausibile immaginare che in futuro la disponibilità di molte valutazioni effettuate sui siti delle Pubbliche Amministrazioni consenta di mettere a confronto sotto-categorie come quella Comuni/Circoscrizioni e Ministeri, e per le quali potrebbe non essere implausibile individuare delle differenze significative che porterebbero, in ultima analisi, alla creazione di norme di taratura più circostanziate. Proprio per questo motivo è importante condurre studi impiegando uno strumento standard che renda possibile il confronto tra diverse valutazioni e diverse tipologie di sito.

²⁰ Portali/Communities, Università, Enti/Pubbliche Amministrazioni, Aziende/ Servizi. Una descrizione di queste macro-categorie (con i relativi valori normativi) è riportata nel capitolo 5.

Il pool di dati impiegato per le analisi riportate in questo manuale è costituito dalle valutazioni effettuate da 1290 utenti su 19 siti. Si tratta di un campione raccolto nel 2013 e, dunque, diverso da quello impiegato per le analisi condotte durante la fase di sviluppo del questionario. I siti Internet riflettono cambiamenti tecnologici e di stile, mentre gli utenti diventano sempre più competenti ed esigenti. Un questionario orientato a stimare la qualità dell'interazione con i siti Internet non può ignorare questi cambiamenti. Per tale motivo, sia per quel che riguarda l'analisi della struttura fattoriale, sia per la realizzazione delle norme di taratura, in questo manuale si farà riferimento -laddove possibile- a dati aggiornati. L'aggiornamento non comporta, in questo caso, l'aggiunta di dati al pool preesistente, ma la sostituzione con dati più recentemente raccolti.

4.2 Struttura Fattoriale e Attendibilità

L'analisi fattoriale²¹ è stata condotta mediante il metodo di estrazione dei minimi quadrati generalizzati. La soluzione a tre fattori è risultata parzialmente correlata e dunque è stata sottoposta a rotazione degli assi mediante il metodo Oblimin. La percentuale di varianza comune spiegata dalla soluzione è risultata pari a 49,28%. Nella tabella successiva vengono riportate le percentuali di varianza spiegate dai singoli fattori

²¹ Per i lettori che non abbiano dimestichezza con la statistica, basti sapere che si tratta di una tecnica che consente di individuare fattori riassuntivi a partire dalla matrice delle correlazioni tra le variabili. In questo caso le variabili sono gli item del questionario e l'analisi fattoriale ha consentito di determinarne un numero inferiore (3) in grado di riassumere quelle originali (19).

dopo la rotazione degli assi e le saturazioni dei fattori sugli item.

M (26,18%)	11. Visitare questo sito è stato facile tanto quanto usare il software che uso di più.	-0,774
	14. In questo sito sono stato sul punto di perdermi.	-0,754
	7. Posso raggiungere l'indice generale degli argomenti con facilità.	-0,705
	1. Esplorando questo sito capisco in ogni momento dove mi trovo.	-0,693
	9. Questo sito è difficile da consultare.	-0,662
	5. Ho provato una sensazione di smarrimento esplorando questo sito.	-0,605
	5. Ho un controllo continuo sulle operazioni possibili in questo sito.	-0,602
	12. In questo sito posso trovare quello che mi interessa senza doverlo esplorare tutto.	-0,598
	3. Fin dall'inizio è chiaro quali sono i contenuti di questo sito.	-0,581
	17. Le informazioni presenti in questo sito sono comprensibili.	-0,443
19. Questo sito presenta lunghe liste di difficile consultazione.	-0,368	

Tabella 4. Saturazione del fattore "Maneggevolezza" nei singoli item.

S (16, 53%)	18. Esplorare questo sito è stata una perdita di tempo.	0,763
	8. Questo sito è inutile con la parvenza di essere utile.	0,761
	4. Questo sito è tanto bello quanto inutile.	0,698
	2. Questo sito ha deluso le mie aspettative.	0,619
	15. Sono riuscito ad ottenere le informazioni/servizi che cercavo.	-0,394
	13. I contenuti di questo sito sono aggiornati.	-0,308

Tabella 5. Saturazione del fattore "Soddisfazione" nei singoli item.

A (6, 57%)	6. La scelta dei colori utilizzabili in questo sito è elegante.	0,687
	10. Gli elementi grafici impiegati sono accattivanti e curati nei particolari.	0,661

Tabella 6. Saturazione del fattore "Attrattiva" nei singoli item.

L'attendibilità dei fattori è stata stimata impiegando la matrice delle covarianze per i punteggi fattoriali. Il livello di attendibilità è risultato soddisfacente per tutte le scale: $M = 0,92$; $S = 0,88$; $A = 0,74$ ²².

²² La numerosità degli item che compongono le tre scale corrisponde alla loro importanza relativa nel definire l'interazione con un sito. Nello sviluppo delle versioni del questionario successive alla prima, si è scelto di mantenere questa disparità derivante dall'analisi fattoriale (piuttosto che pareggiare il numero di item per scala) perché rappresenta più fedelmente il rapporto esistente tra le tre dimensioni. Ciò, ovviamente, ha un effetto sull'attendibilità della scala "A" che, per quanto soddisfacente, risulta ridotta a causa della sua composizione limitata a due soli item.

4.3 Validità

Uno studio sperimentale sulla validità (di costruito e predittiva) dello strumento è stato realizzato impiegando la versione 1.0 per valutare siti sperimentali appositamente realizzati per possedere le caratteristiche in grado di elicitarne gli aspetti di interazione misurati da ogni scala del questionario (Terenzi, Di Nocera e Ferlazzo, 2006). In particolare, la presenza di aiuti di navigazione, la presenza di informazioni necessarie per il completamento delle attività e la gradevolezza estetica sono state manipolate per sottoporre a verifica l'ipotesi che le scale Maneggevolezza, Soddisfazione e Attrattiva mostrassero una specifica sensibilità all'interazione degli utenti con questi aspetti. Ciascuna caratteristica poteva quindi essere presente o assente (variando il numero di scorciatoie, la disponibilità di informazioni e i colori utilizzati), dando origine a tutte le possibili combinazioni di caratteristiche valutate attraverso le scale Maneggevolezza (presente/assente), Soddisfazione (presenza/assenza), Attrattiva (presenza/assenza).

L'esperimento, che ha coinvolto 160 partecipanti, consisteva nell'eseguire alcuni compiti di navigazione impiegando una delle diverse versioni del sito sperimentale all'interno di una condizione di laboratorio controllata. I risultati hanno mostrato valutazioni di usabilità significativamente differenti in funzione della presenza/assenza della caratteristica esaminata, a eccezione della Soddisfazione. Difatti, mentre per le altre due scale è stato possibile creare istanze del sito sperimentale che rendessero difficoltosa la navigazione del sito e risultassero esteticamente poco piacevoli e piatte, la mera indisponibilità di informazioni necessarie a completare il compito non è risultata sufficiente a modulare le valutazioni dei soggetti rispetto

la soddisfazione. Ciò conferma la natura soggettiva della valutazione di usabilità che, in presenza di obiettivi definiti dallo sperimentatore, non ha ragione di esistere.

A differenza dello studio sopra riportato, un compito sperimentale con obiettivi compatibili con le reali esigenze dei partecipanti ha mostrato l'efficacia della scala S nel discriminare il raggiungimento dell'obiettivo. Il grafico successivo mostra i risultati²³ di valutazioni effettuate con Us.E .2.0 da parte di 200 soggetti cui era stato assegnato un compito da eseguire in un sito reale (e di cui erano utenti effettivi o potenziali).

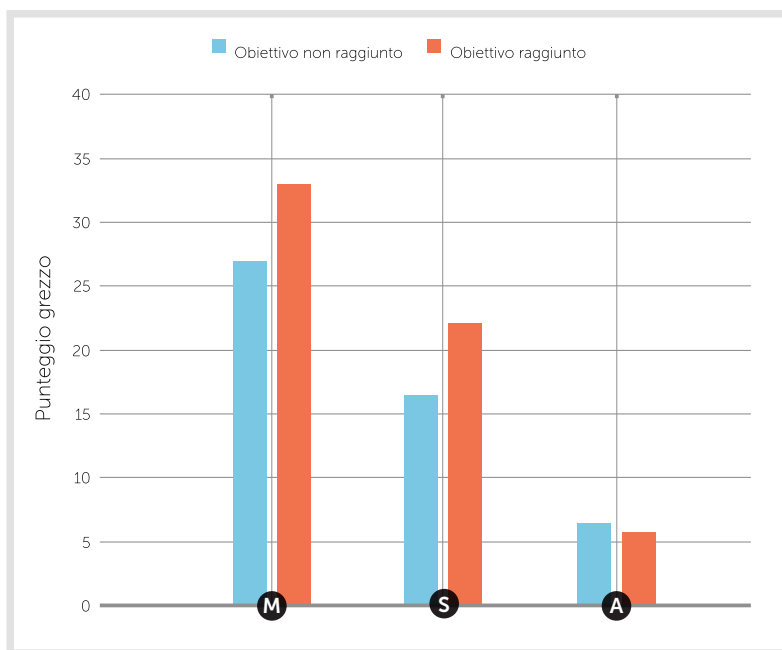


Figura 1. Differenze nella valutazione di usabilità in funzione del raggiungimento dell'obiettivo di navigazione.

²³ Manoscritto in preparazione.

Le valutazioni di usabilità sono significativamente differenti in funzione del successo nel raggiungere l'obiettivo di navigazione. Non sorprendentemente, questa variabile non ha inciso sulla valutazione dell'Attrattiva.

5. Somministrazione, scoring e interpretazione dei punteggi

Nei capitoli precedenti sono state fatte specificazioni e digressioni a beneficio dei lettori che non hanno dimestichezza con il tema dell'usabilità o con la psicomtria. Anche in questo capitolo sarà necessario specificare alcuni aspetti che potrebbero non risultare ovvi a chi, per la prima volta, si trova a impiegare un questionario. In particolare, occorre tener presente che il questionario Us.E. 2.0 deve essere somministrato nella sua interezza, senza modificare le affermazioni che lo compongono. Tutti gli item, difatti, sono indispensabili per calcolare correttamente il punteggio delle tre scale. È consigliabile includere anche le domande che fanno parte della sezione anagrafica, ma ai fini del calcolo dei punteggi solo il genere dei rispondenti è indispensabile²⁴. Relativamente la parte anagrafica, è importante evidenziare come il valutatore interessato a specifici aspetti di composizione del campione sia libero di aggiungere qualunque domanda ritenga utile al fine di segmentare i dati raccolti. Per esempio, un valutatore potrebbe essere interessato a confrontare le stime di usabilità offerte da utenti che si connettono al sito prevalentemente da un computer

²⁴ I punteggi standard devono essere calcolati separatamente per maschi e femmine (vedi capitolo 4).

desktop e quelle di utenti che si connettono prevalentemente impiegando dispositivi mobile. Al momento della stesura di questo manuale, questa domanda non fa parte della sezione anagrafica “standard”, ma ciò non toglie che il ricercatore o il professionista interessato a indagare questo specifico aspetto possa includerla.

5.1 Somministrazione

La somministrazione del questionario può avvenire nella tradizionale forma carta e matita (utile nella ricerca di laboratorio, in cui l’utente è fisicamente presente) oppure on-line. Se si utilizza la somministrazione via web è opportuno assicurarsi che i rispondenti siano effettivamente utenti di quel sito. Si può aumentare la probabilità di questa corrispondenza inserendo un link al questionario ben visibile nella home page del sito da valutare: se il rispondente ha visto il link, evidentemente ha avuto accesso al sito²⁵. Invitare le persone a rispondere al questionario tramite e-mail, annunci su forum e social network può inficiare il campionamento se non si è relativamente sicuri che chi accetta l’invito sia un utente reale o potenziale del sito in esame. Infine, la somministrazione on-line può avvenire impiegando un form appositamente realizzato dal valutatore (per esempio impiegando Google Forms® e strumenti analoghi), oppure, in maniera facilitata, attraverso il sito di supporto al questionario (vedi sezione 5.5).

Somministrazione
via web o carta
e matita?

²⁵ E’ opportuno che il link sia chiaramente visibile e che l’invito a fornire volontariamente una valutazione sia correttamente formulato. In generale, frasi semplici e chiare come “Valuta questo sito” o “Aiutaci a migliorare questo sito” si sono rivelate utili allo scopo.

5.2 Numerosità minima

Come riportato nel capitolo 3, il progetto Us.E. ha prodotto negli anni diverse versioni del questionario. L'obiettivo era, da un lato, ottenere informazioni sempre più accurate sulle dimensioni che sottendono la valutazione di usabilità da parte degli utenti, dall'altro, individuare gli item più rappresentativi delle scale in modo da evitare agli utenti l'onere di rispondere all'iniziale set di 70 item (versione 1.0).

Successivamente il rilascio della versione 2.0, è stato ritenuto opportuno affrontare anche un'altra problematica spesso manifestata dagli utilizzatori del questionario: individuare il numero minimo di utenti da coinvolgere nelle sessioni di valutazione. Il professionista impegnato nella valutazione di usabilità, difatti, spesso si ritrova a dover fornire rapidamente delle indicazioni per la progettazione o ri-progettazione del sito e il tempo necessario per raccogliere dati su numerosi utenti non è quasi mai disponibile. Avere indicazioni chiare rispetto al numero adeguato di protocolli da analizzare è cruciale per l'attività professionale.

Al fine di individuare una numerosità minima che garantisca stime stabili su tutte le scale del questionario, è stata impiegata una procedura di ri-campionamento con re-immissione per costruire le distribuzioni empiriche delle medie di campioni di diversa numerosità (Efron e Tibshirani, 1993). Si tratta di un argomento molto tecnico e che non è necessario conoscere per l'uso del questionario. Cionondimeno, è opportuno riportare (seppur in maniera superficiale) la logica dello studio che ha consentito di definire il numero minimo di utenti oltre il quale è sconsigliabile scendere. In questo studio (Di Nocera et al., 2009)

è stata generata una grande quantità di campioni “fittizi” sulla base di dati reali (un campione di 500 utenti). I campioni creati potevano essere composti di 200, 150, 100, 50, 40, 30, 20, 10 e 5 “utenti” (3000 campioni per ogni numerosità). Sulla base di ognuna di queste numerosità sono quindi state calcolate le medie dei campioni e derivate le relative distribuzioni empiriche. A questo punto è stato possibile confrontare le medie reali di campioni di numerosità pari a 200, 150, 100, 50, 40, 30, 20, 10 e 5 utenti con le corrispondenti distribuzioni empiriche. Se i valori reali ricadevano all’interno dell’intervallo di confidenza (fissato al 95%), il dato veniva considerato “stabile”. La tabella successiva illustra la stabilità delle medie per ogni scala in funzione della numerosità del campione (con “NO” sono stati

	MAN	SOD	ATT
Da 200 a 30 utenti	OK	OK	OK
20 utenti	OK	OK	NO
10 utenti	OK	OK	NO
5 utenti	OK	NO	NO

Tabella 7. Stabilità delle scale in funzione della numerosità del campione impiegato.

indicati i confronti in cui i valori reali ricadevano al di fuori dell'intervallo di confidenza²⁶).

Come è possibile desumere dalla tabella 7, al fine di ottenere una stima stabile nelle tre scale (Maneggevolezza, Soddisfazione, Attrattiva) è necessario somministrare il questionario a un campione che abbia una numerosità minima di 30 utenti. Naturalmente, qualora tra gli scopi della valutazione ci fosse quello di confrontare categorie diverse di utenti, sarebbero necessari campioni di numerosità minima pari a 30 per ognuno dei livelli di quella categoria (per esempio, 30 utenti alla prima visita vs. 30 utenti abituali).

5.3 Norme di taratura

Nel tempo, Us.E. è stato somministrato a un elevato numero di utenti per valutare siti molto diversi tra loro e il punteggio medio ottenuto su ogni scala costituisce oggi una sorta di "àncora" che consente di riferire il punteggio ottenuto in una specifica valutazione a un punteggio normativo che è stato costruito attraverso le precedenti esperienze di valutazione. Per calcolare il punteggio standardizzato (vedi prossimo paragrafo) occorrono tre informazioni: il punteggio ottenuto nella valutazione effettuata, il punteggio medio ottenuto nelle precedenti valutazioni, la deviazione standard dei punteggi ottenuti nelle precedenti valutazioni. Queste ultime due informazioni costituiscono le norme di taratura.

²⁶ L'espressione "intervallo di confidenza" indica un intervallo di valori plausibili per un parametro (in questo caso, la media).

Come riportato nel capitolo 4, le norme di taratura sono suddivise in funzione di quattro tipologie di siti: Portali e Communities, Università, Enti e Pubbliche Amministrazioni, Aziende e Servizi. Inoltre per ogni categoria sono stati individuati valori normativi separatamente per maschi e femmine. La valutazione di usabilità attraverso Us.E. risente delle differenze di genere e occorre quindi riportare il valore ottenuto nella singola valutazione da parte di un utente alle norme specifiche per il suo genere di appartenenza.

Portali e Communities. Rientrano in questa categoria i siti i cui contenuti sono creati o commentati dagli utenti. Questi siti sono frequentati da gruppi di persone che condividono gli stessi interessi sui quali scambiano pareri e opinioni, interagendo in modo continuativo, ponendo (e rispondendo a) domande, nonché condividendo informazioni. La diffusione dei social network (che tutto sommato svolgono le medesime funzioni) non ha ancora determinato un ridimensionamento dell'importanza di questi siti: un recente studio ha riportato come 8 milioni di italiani interagiscono abitualmente in portali e comunità online²⁷.

Università. Rientrano in questa categoria i siti d'Ateneo, delle Facoltà, dei Dipartimenti, delle Biblioteche universitarie, che supportano e veicolano didattica, ricerca e servizi amministrativi. Sono fruiti principalmente da utenti giovani con buone competenze informatiche e determinati ad apprendere l'utilizzo di un sistema che costituisce ormai il modo privilegiato

²⁷ Dati ricavati dalla ricerca "Italia 2.0" che Duepuntozero Doxa conduce dal 2009 sui social media italiani.

per accedere a informazioni ed eseguire procedure (iscrizione ai corsi, prenotazione esami, consultazione del calendario delle attività, ecc.). Questi siti si rivolgono anche al personale docente e non docente che utilizza a vario titolo le funzionalità dei siti accademici.

Enti e Pubbliche Amministrazioni. Rientrano in questa categoria i cosiddetti siti di e-Government, quali: Comuni, Regioni, Ministeri, Enti previdenziali, Istituti nazionali, Aziende sanitarie. Siti molto complessi per la varietà delle informazioni che veicolano e dei servizi che offrono, nonché per l'eterogeneità dell'utenza.

Aziende e Servizi. Rientrano in questa categoria tutti quei siti di e-commerce che consentono l'acquisto di prodotti e servizi, *business-to-business* (B2B), rivolti alle imprese o, B2C (*business-to-consumer*), dedicati ai privati. Quest'ultimo è il mercato più ampio e si rivolge, in Italia, a un'utenza con età media di 37 anni (prevalentemente uomini). Si tratta di un'utenza ancora minoritaria rispetto ai cosiddetti "non-shopper" che si limitano a raccogliere informazioni online su prodotti e brand, ma acquistano attraverso canali tradizionali²⁸.

La tabella successiva riporta i valori normativi per le diverse categorie di sito sopra descritte e separatamente per genere e scala (deviazione standard in parentesi).

Sul sito <http://www.use2.it> è possibile prendere visione delle norme di taratura periodicamente aggiornate.

²⁸ Dati ricavati dall'E-Commerce Report "European Digital Behaviour Study 2013" pubblicato da ContactLab.

FEMMINE	M	S	A
Aziende/servizi	42,50 (9,87)	24,24 (4,14)	6,43 (2,15)
Enti/Pubbliche amministrazioni	40,60 (8,74)	22,62 (4,60)	5,85 (1,89)
Portali/Communities	42,94 (8,16)	24,13 (3,74)	6,63 (1,62)
Università	39,48 (9,82)	22,29 (5,87)	5,33 (1,98)
MASCHI	M	S	A
Aziende/servizi	44,09 (7,72)	25,10 (4,68)	6,99 (1,8)
Enti/Pubbliche amministrazioni	41,62 (9,04)	23,50 (5,23)	6,71 (2,08)
Portali/Communities	42,58 (7,93)	24,57 (4,05)	6,61 (1,80)
Università	35,93 (11,11)	20,79 (6,66)	5,47 (2,20)

Tabella 8. Norme di taratura per le quattro categorie di siti e separatamente per maschi e femmine (Punteggi medi e deviazioni standard).

5.4 Scoring, standardizzazione e interpretazione dei punteggi

Il questionario si presenta con 19 item formulati in maniera positiva o negativa, ai quali verrà data una risposta utilizzando una scala di tipo Likert a 5 punti per esprimere il grado di accordo/disaccordo con le affermazioni.

Lo scoring dei risultati ottenuti ha come base di partenza la matrice delle risposte degli utenti del sito oggetto di valutazione. Per convenzione i soggetti sono rappresentati per riga e le variabili sono allineate in colonna.

Per calcolare i punteggi ottenuti nelle tre dimensioni occorre sommare le risposte (fornite dall'utente in una scala da 1 a 5) a tutte le domande che fanno riferimento a ogni singola dimensione. Alcune di queste affermazioni (item) esprimono la dimensione in esame in forma negativa. Per esempio, "Questo sito è tanto bello quanto inutile" è sì rappresentativo della Soddisfazione, ma in termini negativi. E' una pratica comune nella costruzione di un questionario formulare parte degli item in forma negativa. Ciò è opportuno per evitare che gli utenti rispondano in maniera "automatica" generando quello che viene

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1		sex	age	it1	it2	it3	it4	it5	it6	it7	it8	it9	it10	it11	it12	it13	it14	it15	it16	it17	it18	it19
2	user 01	M	30	1	4	3	1	4	2	2	3	2	4	2	3	2	4	3	5	3	4	4
3	user 02	M	28	4	5	2	3	2	3	2	1	5	4	2	2	5	5	3	4	2	5	1
4	user 03	F	55	1	5	3	2	1	5	5	4	2	5	1	1	1	5	4	2	3	3	2
5	user 04	F	21	4	1	4	3	1	3	3	5	2	3	1	3	2	2	4	5	2	2	3
6	user 05	F	33	4	2	1	3	3	3	5	2	3	5	4	4	1	2	3	1	3	2	5
7	user 06	F	24	3	5	1	1	2	2	2	4	1	1	2	1	5	1	5	4	1	5	5
8	user 07	M	19	2	3	2	4	4	3	2	5	2	2	4	3	5	2	5	1	5	1	2
9	user 08	M	27	5	1	2	1	1	2	2	2	5	2	3	1	4	5	3	3	2	1	4
10	user 09	M	54	5	2	1	5	4	3	2	1	4	3	3	4	1	4	3	1	4	1	4
11	user 10	F	34	5	1	4	5	1	4	4	2	3	3	1	4	1	5	1	2	5	4	2
12	user 11	M	40	5	1	2	3	5	5	1	1	4	3	2	5	2	3	4	4	4	5	1
13	user 12	F	35	5	2	3	4	5	5	5	4	1	2	4	1	4	5	1	5	4	3	3
14	user 13	M	22	1	1	5	1	2	1	3	3	3	5	2	5	2	5	1	1	2	5	4
15	user 14	M	34	3	2	1	1	3	3	4	5	3	4	2	4	1	1	2	3	4	2	2
16	user 15	M	18	4	1	5	1	3	4	3	2	1	5	4	1	3	1	2	5	1	4	2
17	user 16	F	44	3	5	3	3	1	3	3	2	3	5	5	1	1	2	1	2	4	5	2
18	user 17	F	31	5	4	3	5	2	5	1	4	2	1	1	3	4	5	1	4	3	3	1
19	user 18	F	56	4	5	4	5	3	4	2	1	2	1	2	5	3	3	3	1	5	5	4
20	user 19	F	23	4	1	5	1	2	1	2	3	2	2	4	4	1	5	3	4	4	5	5
21	user 20	F	25	4	5	5	4	1	1	3	2	3	1	1	3	4	3	1	4	1	2	3
22	user 21	M	41	3	3	5	1	1	5	1	1	3	4	1	1	3	4	2	1	1	2	3
23	user 22	M	38	5	3	2	5	2	2	5	2	5	1	3	2	2	3	1	2	1	4	5
24	user 23	F	27	3	3	2	1	2	1	3	4	5	1	1	3	5	1	3	2	3	5	4
25	user 24	M	22	2	3	5	3	2	3	5	5	4	1	3	3	5	2	4	5	4	3	3
26	user 25	M	55	5	4	5	1	3	5	2	4	5	3	3	3	1	1	2	5	4	1	4

Figura 2. Esempio di matrice di dati U.S.E.

comunemente definito “response set”. Prima di sommare questi dati occorre, dunque, ricodificare le risposte. Il procedimento è semplice: basta sottrarre il punteggio ottenuto in quell’item a 6. Il risultato sarà la perfetta inversione della scala di risposta²⁹. Di seguito vengono riportati i (semplici) calcoli da eseguire per ottenere i punteggi alle tre dimensioni.

²⁹ Sottraendo a 6 il punteggio fornito dall’utente, l’item verrà ricodificato, ottenendo valori speculari. In pratica, “5” diventerà “1”, “4” diventerà “2” e così via. In tal modo la risposta “5” (“Assolutamente d’accordo”) a un item che esprime negativamente il costrutto (per esempio, “Ho provato una sensazione di smarrimento in questo sito”) diventerà “1” (“Assolutamente in disaccordo”), riportando in tal modo l’opinione dell’utente in rapporto alla sua formulazione positiva. In altre parole, è come se l’item fosse stato formulato “Non ho provato una sensazione di smarrimento in questo sito”.

$$M = it1 + it3 + (6 - it5) + it7 + (6 - it9) + it11 + it12 + (6 - it14) + it16 + it17 + (6 - it19)$$

$$S = (6 - it2) + (6 - it4) + (6 - it8) + it13 + it15 + (6 - it18)$$

$$A = it6 + it10$$

I risultati di queste operazioni costituiscono il cosiddetto "punteggio grezzo" ottenuto in ogni singola dimensione. Il punteggio grezzo di una scala (quello ottenuto semplicemente sommando le risposte agli item opportunamente ricodificati) è utilizzabile solo per operare confronti. Per esempio, un valutatore potrebbe voler confrontare il punteggio ottenuto alle tre scale di Us.E. 2.0 prima e dopo un processo di redesign, oppure per mettere a confronto i siti di due o più aziende competitor, o -ancora- mettere a confronto le valutazioni di due o più gruppi di utenti rispetto al medesimo sito (per esempio, si potrebbe voler confrontare la valutazione di usabilità di utenti esperti e occasionali). In questi casi, il questionario viene impiegato come strumento per ottenere informazioni utili a confrontare diverse condizioni e non per stimare l'usabilità del sito in rapporto alla categoria cui esso appartiene.

Per acquisire un significato "assoluto", i punteggi relativi ogni singola valutazione devono essere messi in relazione a quelli ottenuti su siti analoghi (cioè appartenenti alla medesima categoria). Questa procedura è detta "standardizzazione" e si effettua sottraendo il punteggio ottenuto dal sito oggetto di valutazione al punteggio medio del campione di taratura (il nostro campione di riferimento) e dividendo per la deviazione standard.

$$Z = \frac{X - \bar{X}}{\sigma}$$

Il punteggio così ottenuto (punteggio “z”) rappresenta la posizione di quel sito per quella specifica dimensione rispetto ai siti della stessa categoria³⁰. In tutti gli studi condotti finora, maschi e femmine forniscono stime significativamente differenti sulle tre scale. Queste differenze possono essere attribuite all’expertise, agli obiettivi di navigazione e alle preferenze estetiche. In ogni caso, data questa differenza, ciò comporta che si debba standardizzare il punteggio del singolo utente impiegando l’opportuna norma di taratura in funzione del genere e poi procedere a calcolare il valore medio per il sito.

Ipotizziamo che un utente maschio abbia ottenuto un punteggio pari a 23 nella scala “M”, che la media del campione di riferimento per quella scala sia di 44,09 e che la deviazione standard del campione sia di 7,72. A questo punto, applicando la formula sopra citata, avremo:

$$z = \frac{23 - 44,09}{7,72} = 2,73$$

e così via per le altre scale e per ogni singolo utente. La media di tutti gli utenti determinerà infine la valutazione del sito in esame.

³⁰ La standardizzazione è una procedura necessaria per dare significato ai punteggi ottenuti attraverso un questionario. Un test, difatti, è uno strumento di misura e necessita di una scala di riferimento. Un punteggio pari a 10 non è dotato di significato di per sé, piuttosto assume significato se sappiamo quanto è distante dal punteggio ottenuto impiegando quel test in altre valutazioni. Analogamente, non possiamo definire una persona “alta” o “bassa” senza sapere quale sia la statura media di riferimento.

Il passo successivo consiste nel disegnare il grafico a barre corrispondente ai punteggi z ottenuti per ciascuna scala, ottenendo, così, il profilo del sito.

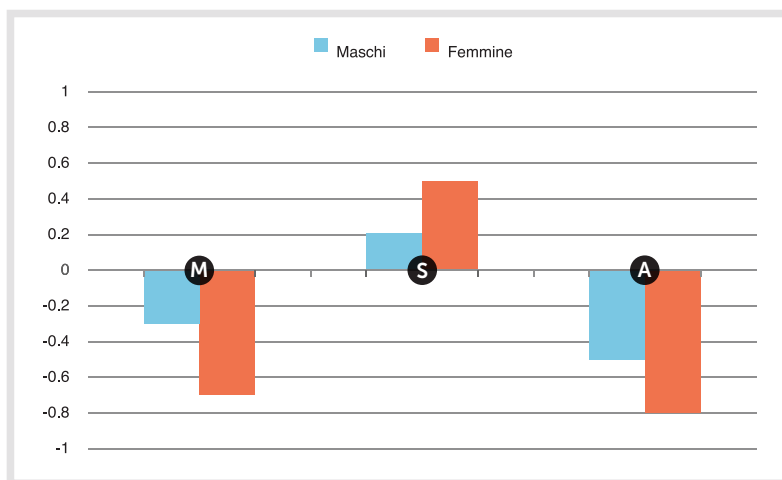


Figura 2. Grafico a barre esemplificativo per riassumere il risultato della valutazione condotta.

Punteggi positivi indicano che, per il sito in esame, quella dimensione è meglio rappresentata rispetto ai siti di quella categoria. Ovviamente, punteggi negativi indicano che la dimensione è al di sotto dello standard dei siti di quella categoria. Punteggi superiori a 1 e inferiori a -1 indicano uno scostamento di oltre una deviazione standard³¹ dalla media di riferimento.

I profili così ottenuti consentono l'interpretazione dei punteggi e permettono di suggerire indicazioni per il miglioramento degli aspetti critici del sito limitatamente alle scale che mostrano valori di z inferiori alla media. Il valore di Us.E. 2.0

³¹ La deviazione standard è uno dei modi per esprimere la dispersione dei dati intorno ad un indice di posizione.

risiede soprattutto in questa capacità di fornire indicazioni circostanziate in maniera economica e veloce, perché ottenute attraverso la somministrazione di soli 19 item a un campione minimo di 30 utenti.

5.5 Sito di supporto al questionario

Fin dalla sua nascita, il questionario Us.E. è sempre stato accompagnato da un sito di supporto (oggi disponibile all'indirizzo <http://www.use2.it>) in costante aggiornamento e che costituisce uno strumento indispensabile per rendere più efficiente il lavoro del valutatore.

Al momento della stesura di questo manuale, il sito consente di:

1. creare un'istanza personalizzata del questionario per la somministrazione on-line;
2. raccogliere i dati dei rispondenti su un server;
3. standardizzare automaticamente i punteggi;
4. creare il grafico riassuntivo della valutazione;
5. esportare i dati in formato CSV per procedere a ulteriori analisi dei dati e per la lettura dei commenti lasciati dagli utenti.

Inoltre, è presente un servizio di supporto tecnico gratuito per il corretto uso della piattaforma e del questionario.

Per usufruire dei servizi offerti, occorre effettuare una procedura di registrazione gratuita. L'iscrizione comporta l'adesione al programma di aggiornamento delle norme di taratura: i dati raccolti, pur restando di proprietà del valutatore, diventano

automaticamente candidabili per essere inseriti nel *pool* dei siti valutati nelle diverse categorie e che collettivamente costituiscono i dati normativi. I dati vengono ovviamente impiegati in forma anonima, ma ogni valutatore può scegliere, qualora sia qualificato a farlo, di consentire l'inserimento del sito valutato come "testimonial" del progetto.

Dopo aver effettuato il login al sito use2.it è possibile gestire il proprio account e le proprie valutazioni. Per creare una sessione di valutazione, occorre inserire il nome e l'indirizzo del sito che si vuole valutare e la tipologia di sito da valutare. Le tipologie di sito finora supportate (data la disponibilità di norme di taratura adeguate) sono:

1. Azienda / Servizio
2. Ente / Pubblica Amministrazione
3. Portale / Community
4. Università

In futuro, altre categorie di sito verranno introdotte compatibilmente con l'aggiornamento delle norme di taratura. Nella pagina "i miei siti" vengono riportate le valutazioni precedentemente effettuate e quelle in corso. In particolare, per ogni valutazione vengono riportate le seguenti informazioni:

- il link per accedere alla propria istanza personalizzata del questionario;
- il numero di rispondenti;
- un link ai dati (scaricabili in formato .csv);
- un link al grafico riassuntivo della valutazione (scaricabile in formato .png).

Ogni valutazione avrà dunque il suo specifico link a una copia del questionario da somministrare on-line. E' molto importante che non si faccia confusione tra i link associati ai diversi siti da valutare. Difatti, ogni link rimanda a una copia personalizzata di Us.E. 2.0, la quale riporta al suo interno il nome del sito da valutare, nonché (nella stringa del link) il codice unico associato a quella valutazione. Dunque, i rispondenti si collegheranno a una specifica copia del questionario associata al sito da valutare e i dati raccolti saranno attribuiti alla specifica valutazione di cui fanno parte.

Il link può essere inviato via email, diffuso sui social network e sui forum, o -più opportunamente- può essere inserito nella home page del sito da valutare. Il modo corretto di impiegare il questionario Us.E. 2.0 consiste nel somministrarlo agli utenti effettivi del sito in esame. Collocando il link al questionario nella home page del sito da valutare (eventualmente etichettandolo come "valuta questo sito", oppure "aiutaci a migliorare questo sito"), si ottiene una visibilità del link limitata agli utenti che si sono effettivamente collegati al sito.

Ciononostante, possono esistere diverse buone ragioni per inviare il link via email o diffonderlo su piattaforme di condivisione (social, forum, ecc.). In alcuni casi, difatti, potremmo voler valutare un sito ancora sperimentale e per il quale abbiamo già identificato un bacino di utenza che decidiamo di raggiungere direttamente con un invito esplicito a partecipare alla valutazione. Similmente, potremmo decidere di non far compilare il questionario on-line e di invitare un campione di utenti a visitare il sito in condizioni controllate di laboratorio (per esempio, assegnando dei compiti di navigazione specifici) e impiegare il questionario nella sua forma cartacea. Anche

questa è disponibile all'indirizzo <http://www.use2.it>, in formato PDF e liberamente scaricabile.

A questo proposito è importante sottolineare che, in tutte quelle circostanze in cui si sceglierà il campione di utenti che risponderà al questionario, occorrerà prestare particolare attenzione alla selezione dei partecipanti. E' necessario che i partecipanti siano utenti reali o almeno potenziali del sito in esame e che la compilazione del questionario sia subordinata all'effettiva esplorazione del sito.

Appendice 1: Come scrivere un report di valutazione

Per il professionista impegnato nella valutazione di usabilità, una volta effettuata la valutazione tramite Us.E. e aver calcolato i punteggi standard in funzione delle opportune norme di taratura, la stesura di un report di valutazione costituisce la vera e propria fase di “delivery” dei risultati al cliente che ha commissionato lo studio.

Un rapporto tecnico non deve necessariamente avere una struttura standard. Molto dipende dalla natura della consulenza effettuata e dalle abilità del singolo professionista. Tuttavia, alcuni elementi sono ineludibili. Di seguito viene riportato un elenco di queste voci, una sorta di checklist per la sua stesura:

- una descrizione dello strumento impiegato che includa il riferimento bibliografico al presente manuale;
- una descrizione del campione di utenti (numerosità, genere, età media, livello di esperienza, ecc.);
- punteggi standard nelle tre scale;
- riferimento alla norma di taratura impiegata;
- grafico dei punteggi standard nelle tre scale (eventualmente più d'uno se la segmentazione dei risultati è giustificata da differenze non marginali tra categorie di utenti);
- discussione dei risultati;
- proposte di intervento.

Se presenti e significativi, possono essere introdotti nel report riferimenti ai commenti lasciati dagli utenti, che spesso si rivelano molto utili per interpretare i risultati della valutazione. Naturalmente, possono essere inseriti anche i risultati di analisi statistiche eventualmente condotte e riferimenti bibliografici a

studi che il valutatore intende citare per sostenere le proprie considerazioni.

Allo scopo di fornire indicazioni operative su come realizzare un report "di minima", di seguito viene riportato un esempio realizzato impiegando dati fittizi³². Si tratta solo di una linea guida ed è sconsigliabile imitarne pedissequamente la forma e i contenuti: come già riportato, l'esperienza del valutatore, le specificità della consulenza, come pure le necessità del committente potranno influenzare la forma e il dettaglio del report stesso³³.

Usability Evaluation 2.0 - Report di valutazione

L'usabilità del sito oggetto della presente valutazione³⁴ è stata stimata mediante il questionario Usability Evaluation (Us.E.). Nella sua versione corrente (Us.E. 2.0: Di Nocera, 2013) il questionario è composto di 19 affermazioni rispetto alle quali l'utente esprime il proprio grado di accordo/disaccordo attraverso una scala di tipo Likert a 5 punti (da "assolutamente falso" ad "assolutamente vero"). La stima dell'usabilità viene espressa attraverso tre dimensioni: Maneggevolezza (scala "M"), Soddisfazione (scala "S") e Attrattiva (scala "A").

³² In questo esempio è stata simulata la valutazione di un sito commerciale (categoria di riferimento: Aziende / Servizi).

³³ Per approfondire il tema della reportistica, si consiglia la lettura della ISO/IEC 25062:2006 "Common Industry Format (CIF) for usability test reports".

³⁴ In un report reale è opportuno inserire qui il nome del sito e la sua URL.

- La scala "M" fornisce indicazioni relativamente l'interazione con le caratteristiche strutturali del sito Internet in esame. Punteggi al di sotto della media in questa scala suggeriscono la necessità di affrontare cambiamenti strutturali, come la riorganizzazione della gerarchia dei contenuti, il riposizionamento degli elementi all'interno delle pagine, l'aggiunta o la modifica di un menu di navigazione.
- La scala "S" fornisce indicazioni relativamente l'utilità percepita del sito ed esprime quindi quanto l'utente consideri vantaggioso accedere al sito Internet in esame per soddisfare le proprie esigenze. Punteggi inferiori alla media in questa scala potrebbero indicare che il sito non è stato realizzato tenendo conto del suo bacino di utenza.
- La scala "A" fornisce informazioni sulle impressioni di gradevolezza suscitate dal sito (o usabilità apparente). Punteggi inferiori alla media potrebbero indicare la necessità di un re-styling, l'aggiunta di funzionalità e una maggiore cura nei particolari.

Campione

Sono state raccolte le valutazioni di 100 utenti (41 maschi, 59 femmine, età media = 28,64 dev.st = 10,46)³⁵ con il seguente

³⁵ Potrebbe essere opportuno confrontare le caratteristiche del campione che ha risposto al questionario con quelle del settore di riferimento: in questo caso la tipologia di utenti che abitualmente fa acquisti on-line.

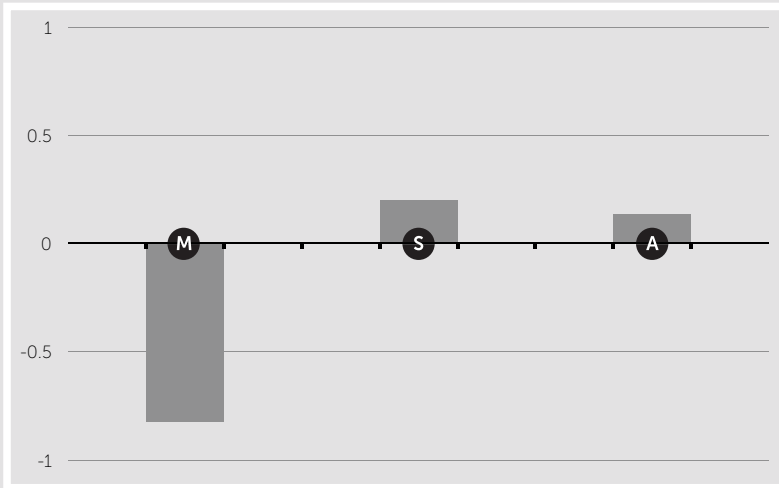
livello di scolarizzazione: Licenza media = 1%; Maturità = 53%; Laurea Triennale = 32%; Laurea Magistrale (o a ciclo unico) = 12%; Dottorato o Master = 2%. Tutti i partecipanti hanno riportato di usare Internet quotidianamente e l'85% dei rispondenti ha dichiarato di aver visitato il sito in precedenza³⁶.

Analisi dei dati

I punteggi ottenuti nelle tre scale sono stati standardizzati impiegando le norme di taratura per Aziende e Servizi fornite dal sito use2.it. I punteggi così ottenuti (punteggi "z") rappresentano il livello di usabilità del sito oggetto della valutazione rispetto alla media dei siti della stessa categoria. Il grafico successivo mostra il profilo di usabilità del sito oggetto della presente valutazione: lo zero rappresenta il punteggio di usabilità medio dei siti commerciali e il range da +1 a -1 rappresenta la variazione pari a una deviazione standard.

Come è possibile evincere dal grafico, gli utenti hanno espresso una valutazione negativa dell'interazione con gli aspetti strutturali del sito (Scala M=-0,83), mentre risultano valutati positivamente -malgrado non si discostino molto dalla media di riferimento- l'utilità percepita (Scala S=0,20) e gli aspetti estetici

³⁶ Strumenti come Google Analytics© sono utilissimi per integrare questi dati e dare significato a eventuali differenze tra categorie di utenti o per selezionare uno specifico sotto-campione. Per esempio, è possibile che il "conversion rate" o il "bouncing" di un sito siano sensibilmente diversi in funzione di alcune caratteristiche demografiche o in base al tipo di dispositivo utilizzato.



del sito (Scala A=0,14). Non sono emerse differenze sostanziali tra le opinioni espresse dai diversi gruppi di utenti (maschi vs. femmine, visitatori nuovi vs. visitatori di ritorno).

Questi risultati sono coerenti con i commenti spontaneamente lasciati dagli utenti³⁷, che confermano le difficoltà di navigazione connesse all'interfaccia, con particolare riferimento alle etichette impiegate nel sito (link, voci di menu) e alla sequenza di operazioni necessarie per effettuare un acquisto.

Conclusioni

I risultati di questa valutazione hanno consentito di identificare nell'interazione con gli aspetti strutturali del sito l'elemento critico da affrontare per migliorare l'usabilità del sito. In

³⁷ Il testo integrale dei commenti potrebbe utilemente essere riportato come appendice del report di valutazione.

particolare, gli utenti trovano le etichette impiegate poco rappresentative e la gerarchia tra i vari elementi non sempre chiara. Queste criticità potranno essere approfondite e risolte impiegando un campione rappresentativo di utenti del sito nelle seguenti attività di analisi e intervento:

- test di usabilità per approfondire le ragioni delle criticità (per esempio, perché l'utente non vede, non comprende, non trova quello che sta cercando) e individuare le aree più carenti (per esempio, sistema di navigazione globale, locale o contestuale);
- riorganizzazione dei contenuti e delle relative etichette mediante l'impiego della tecnica del Card Sorting (Spencer, 2009) che, consentendo di anticipare le modalità di categorizzazione degli utenti, permette di sviluppare strutture che massimizzino la trovabilità dei contenuti.

Riferimenti bibliografici

Di Nocera, F. (2013). *Usability Evaluation 2.0: Una descrizione (s)oggettiva dell'usabilità*. Roma: Ergoproject.

Spencer, D. (2009). *Card Sorting: Designing Usable Categories*. New York: Rosenfeld Media.

Appendice 2: Us.E. 2.0 - English Version

Anche se è inclusa nella roadmap di sviluppo, al momento della stesura di questo manuale non esiste un vero e proprio adattamento di Us.E. 2.0 alla lingua inglese. Per venire a un adattamento comporta: 1) la somministrazione degli item tradotti a un campione di adeguata numerosità, 2) l'esame delle proprietà psicometriche della forma inglese del questionario e 3) la creazione di norme di taratura specifiche. Nel frattempo, e con le dovute cautele, il valutatore interessato potrà impiegare la seguente traduzione degli item ottenuta attraverso una procedura di back-translation³⁸. La scala di risposta sarà, naturalmente, da "strongly agree" a "strongly disagree".

Se Us.E. fosse un'applicazione software, questa versione inglese sarebbe indicata come una "experimental feature" e sarebbe corredata di ogni genere di avvertimento per declinare la responsabilità dell'autore. Il valutatore sia consapevole del fatto che questa versione non può essere impiegata con le norme di taratura italiane, che non esistono ancora norme specifiche per la lingua inglese e che questa traduzione potrebbe in futuro non mostrare le medesime proprietà psicometriche della versione italiana.

³⁸ La procedura translation/back-translation consiste nel far tradurre a due distinte persone bilingue la versione del testo dall'italiano all'inglese e quest'ultima di nuovo in italiano (senza conoscere il testo originale). In tal modo è possibile confrontare la versione originale e quella ottenuta dalla back-translation e valutare la qualità della traduzione. La procedura è iterativa e ha termine quando le due versioni coincidono (o, più realisticamente, le differenze sono accettabili).

1. While exploring this website I always knew where I was
2. This website did not meet my expectations
3. The contents of this website were clear from the beginning
4. This site is as pretty as it is useless
5. I felt disoriented while exploring this website
6. The choice of the colors used in this website is smart
7. I can easily reach the main menu
8. This website is useless while pretending to be useful
9. It is difficult to browse this website
10. The graphic used in this website are catchy and detailed
11. Visiting this website was as easy as using the software application I use the most
12. In this website I can find what I'm looking for without having to explore it all
13. The contents of this website are updated
14. In this website I found myself on the point of getting lost
15. I managed to obtain the information/service that I was looking for
16. I always feel in control of the operations that are allowed in this website
17. The information presented in this website is understandable
18. Exploring this website was a waste of time
19. This website is made up of long lists that are difficult to examine

Tabella 9. Traduzione inglese degli item di Us.E. 2.0

Us.E. è uno strumento creato per la comunità e supportato dalla comunità. Si invitano i ricercatori e i professionisti interessati alla versione in lingua inglese a condividere con il team di sviluppo i dati raccolti con questi item, così da poter pervenire a una versione stabile da impiegare per la valutazione di siti con utenza internazionale.

Riferimenti bibliografici

- Bennet, J.L. (1979), The commercial impact of usability in Interactive systems, in Shackel B. (Ed.), *Man/Computer Communication*, Vol. 2 (pp. 1-17), Maidenhead: Infotech International.
- Boscarol, M. (2003). *Ecologia dei Siti Web*. Milano: Tecniche Nuove.
- Brooke, J. (1996). SUS: A "quick and dirty" usability scale. In P.W. Jordan, B. Thomas, B.A. Weerdmeester, & A.L. McClelland (Eds.), *Usability Evaluation in Industry* (pp. 189-194). London: Taylor and Francis.
- Card, S.K., Moran, T.P., & Newell, A. (1983). *The Psychology of Human-Computer Interaction*. Hillsdale: Lawrence Erlbaum Associates.
- Chin, J.P., Diehl, V.A., & Norman, K.L. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. *Proceedings of SIGCHI '88* (pp. 213-218). New York: ACM/SIGCHI.
- Di Nocera, F., Coletta, S., Ferlazzo, F., & Renzi, P. (2009). Numerosità del campione e stabilità della stima di usabilità ottenuta mediante valutazioni soggettive. *Giornale Italiano di Psicologia*, 3, 695-704.
- Di Nocera, F., Coletta, S., Ferlazzo, F., & Renzi, P. (2009). Numerosità del campione e stabilità della stima di usabilità ottenuta mediante valutazioni soggettive. *Giornale Italiano di Psicologia*, 3, 695-704.

- Di Nocera, F., Ferlazzo, & Renzi, P. (1999). Us.E. 1.0: costruzione e validazione di uno strumento in lingua italiana per valutare l'usabilità dei siti internet. In M.F. Costabile e F.Paternò (a cura di), *Atti di HCITALY '99: Giornata italiana su Human-Computer Interaction*. Rapporto CNUCE-B4-1999-003.
- Di Nocera, F., Ferlazzo, F., & Renzi, P. (2003). L'usabilità a quattro dimensioni. *Ricerche di Psicologia*, 26(4), 83-104.
- Di Nocera, F., Terenzi, M., & Ferlazzo, F. (2007). Misurare l'Information Handling: Un contributo Sperimentale della scala "Manegevolezza" di Us.E. 1.0. *Giornale Italiano di Psicologia*, 24(1), 223-234.
- Efron, B., Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall.
- Johnson, J. (2010). *Designing with the Mind in Mind: Simple Guide to Understanding User Interface Design Rules*. Waltham: Morgan Kaufmann.
- Kirakowski, J., Claridge, N., & Whitehand, R. (1998). Human Centered Measures of Success in Web Site Design. *Proceedings of the 4th Conference on Human Factors & the Web*, Baskin Ridge (NJ), June 5.
- Kurosu, M., Kashimura, K. (1995). Apparent Usability vs Inherent Usability: Experimental Analysis on the Determinants of the Apparent Usability. *Conference Companion on Human Factors in Computing Systems* (pp. 292-293). New York: ACM Press.
- Raskin, J. (2000). *The Humane Interface: New Directions for*

Designing Interactive Systems. Reading: Addison-Wesley.

Terenzi, M., Di Nocera, F., & Ferlazzo, F. (2006). Firmitas, Utilitas, Venustas: Assesding the Validity of the Usability Evaluation Questionnaire. In D. de Waard, K.A. Brookhuis, A. Toffetti (Eds.), *Developments in Human Factors in Transportation, Design, and Evaluation* (pp. 249-253). Maastricht, The Netherlands: Shaker Publishing.

Travis, D. (2006). Red route usability: *The key user journeys with your web site*. Retrieved from <http://www.userfocus.co.uk/articles/redroutes.html>

Avere una presenza on-line costituisce oggi un importante indicatore di credibilità e successo per le organizzazioni. Essere privi di un sito Internet rappresenta uno svantaggio competitivo, ma allo stesso tempo una visibilità on-line di scarsa qualità può danneggiare aziende ed enti che vogliono offrire informazioni, servizi o prodotti ai propri clienti/utenti attraverso Internet.

Questo manuale costituisce una descrizione esaustiva di uno strumento valido e affidabile specificatamente realizzato per fornire indicazioni rapide e continue sulla qualità dell'interazione tra utenti e siti Internet. Usability Evaluation (Us.E.) è un questionario multidimensionale che consente di effettuare una diagnosi rapida dell'usabilità, individuando eventuali criticità per le quali possa essere necessario approfondire l'analisi per poi intervenire attraverso il re-design.

Il manuale, corredato di informazioni sulla somministrazione del questionario, lo scoring e la stesura di un report, costituisce un ausilio indispensabile per chiunque sia impegnato nella progettazione e realizzazione di siti Internet.

(Foto di Valentina Fontanella)



Francesco Di Nocera *insegna Ergonomia Cognitiva e Processi Decisionali presso la Facoltà di Medicina e Psicologia di Sapienza Università di Roma. La sua attività di ricerca è prevalentemente indirizzata allo studio dell'interazione tra individui e tecnologia. Autore di numerose pubblicazioni su questo argomento, è membro del Gruppo di Lavoro per l'Usabilità (GLU) del Dipartimento della Funzione Pubblica.*