



Università degli Studi di Milano Bicocca

Scuola di Scienze

Dipartimento di Informatica, Sistemistica e Comunicazione

Corso di laurea in Informatica

Sviluppo e valutazione di visualizzazioni per l'espressione intuitiva di probabilità e rischi

Relatore: *Federico Cabitza*

Co-relatore: *Andrea Campagner*

Relazione della prova finale di:

Laura Nesossi

Matricola 817166

Anno Accademico 2018-2019

*A Davide, senza il quale questa tesi sarebbe ancora
un mucchio di ansia e pagine bianche.*

Ad Ila, Lodo, Jacopo, Pezz e Ste, per il supporto tecnico e morale.

A Vittorio, per avermi sopportata e aver svolto l'analisi dei questionari.

A Paola, per l'insperata pazienza nell'eseguire troppi test utente.

Al mio relatore, per il sostegno e l'entusiasmo,

che mi hanno fatto credere un po' più in me stessa.

A Bonbon, per aver assistito paziente alla scrittura e riscrittura della tesi.

Al rosa, per rendere tutto un po' più bello.

A me stessa, perché questo sia solo l'inizio di tante belle cose.

Indice

- 1.** Introduzione e obiettivi
- 2.** Sviluppo dell'ambiente web
 - 2.1.** Creazione della pagina web
 - 2.2.** Definizione e sviluppo dei singoli effetti
 - 2.2.1.** Rimozione random di pixel
 - 2.2.2.** Rimozione ordinata di pixel
 - 2.2.3.** Trasparenza
 - 2.2.4.** Scambio random di pixel
 - 2.2.5.** Blur
- 3.** Test utente
 - 3.1.** Definizione degli obiettivi
 - 3.2.** Definizione degli utenti
- 4.** Risultati test utente
 - 4.1.** Efficacia: Checklist task
 - 4.1.1.** Efficacia assoluta
 - 4.1.2.** Efficacia relativa
 - 4.2.** Efficacia: Slider task
 - 4.2.1.** Efficacia assoluta
 - 4.2.2.** Efficacia relativa
 - 4.3.** Efficienza
- 5.** Questionario di valutazione
 - 5.1.** Distribuzione risposte
 - 5.2.** Analisi per scala
 - 5.2.1.** Attrattività
 - 5.2.2.** Apprendibilità
 - 5.2.3.** Originalità
 - 5.2.4.** Stimolazione
 - 5.2.5.** Controllabilità
 - 5.2.6.** Efficienza
 - 5.3.** Confronto fra scale
 - 5.4.** Net Promoter Score
- 6.** Conclusioni
- 7.** Appendice
- 8.** Biografia

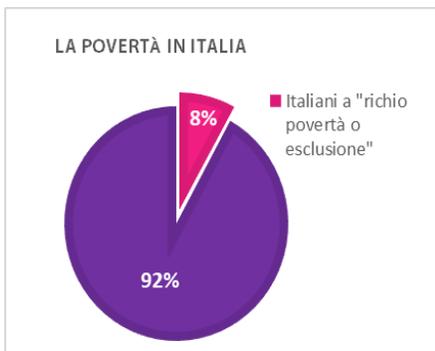
1. INTRODUZIONE E OBIETTIVI

Nell'esecuzione del proprio lavoro i medici si avvalgono del supporto di diverse tecnologie, che restituiscono dati numerici e informazioni allo stato grezzo. La capacità di filtrare e comprendere i diversi tipi di informazione, dunque, assume un ruolo di primaria importanza. Ma i dati grezzi non sono sempre facili da analizzare, in particolare per chi non sia un analista.

Risulta per questo importante adottare metodologie di rappresentazione dei dati che sappiano staccarsi dal mero valore numerico, e fornirne immediatamente l'interpretazione, consentendo al professionista di concentrarsi su diagnosi e rapporto col paziente, più che sull'interpretazione dei dati.

In particolare la rappresentazione visiva ha assunto un ruolo importante nella storia della comunicazione, dalle immagini sui bassorilievi degli antichi egizi alle raffigurazioni sul funzionamento del cosmo di filosofi e astrologi, fino a diventare un vero e proprio strumento utile a comunicare informazioni.

La prima infografica riconosciuta come tale venne pubblicata nel 1626 da un astronomo tedesco, Christoph Scheiner, con l'obiettivo di rappresentare i modelli seguiti dalla rotazione del sole¹.



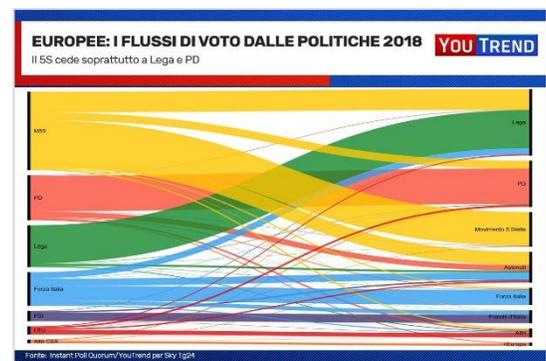
1 Esempio di infografica nella quale l'attenzione cade sul valore numerico

Dal XIX secolo in poi lo strumento dell'infografica si è sempre più diffuso, anche nella comunicazione di massa, affermando la sua natura di mezzo facile ed intuitivo per la comunicazione dei dati. Ci sono però delle problematiche che vengono a galla in contesti d'uso particolari, come può essere quello medico.

In primo luogo, alcune infografiche risultano eccessivamente legate al numero. Ciò implica che anziché fornirci immediatamente l'informazione, la nostra mente si rifà al valore numerico, per poi assumerne delle conclusioni. Questo può risultare problematico anche perché non tutti tendono a valutare allo stesso modo valori numerici uguali: una probabilità di ammalarsi del 5% può sembrare alta a qualcuno, ma bassa per qualcun altro, anche in funzione di diversi retaggi culturali.

Altre infografiche, invece, risultano poco chiare ad una prima occhiata, e richiedono un'analisi attenta per rendere l'informazione comprensibile. Può diventare un problema in situazioni in cui è necessaria chiarezza e rapidità, o in cui è richiesto di scorrere grandi quantità di visualizzazioni diverse.

L'obiettivo di questa tesi è quello di analizzare una nuova modalità di visualizzazione grafica di percentuali, che punta ad essere intuitiva e rapida da comprendere, ma senza essere palesemente collegabile ad un valore numerico. L'idea è quella di creare dei filtri che riducano la chiarezza dell'immagine a cui vengono



2 Esempio di infografica che necessita di un certo sforzo cognitivo per essere compresa²

applicati, in maniera proporzionale al ridursi della percentuale che essa vuole rappresentare. Dunque, dato un valore percentuale del 100%, l'infografica che lo rappresenta sarà perfettamente nitida. La chiarezza diminuirà al diminuire della percentuale, fino a diventare irriconoscibile al raggiungimento dello 0%.

Dopo aver individuato dei filtri che potrebbero essere adeguati, sarà necessario testarne la comprensibilità, con l'aiuto di utenti reali.

2. SVILUPPO DELL'AMBIENTE WEB

Per realizzare le visualizzazioni grafiche è stato sviluppato un sito che rendesse possibile modificare un'immagine in modo che rappresentasse una percentuale scelta, mediante un effetto selezionato. Le tecnologie utilizzate per la costruzione dell'applicazione web sono state HTML, CSS e JavaScript.

Per la parte grafica ci si è avvalsi del supporto della libreria *Bootstrap*³, una raccolta di strumenti liberi per la creazione di siti e applicazioni per il Web. Trattandosi di codice già testato e in grado di semplificare la costruzione di un sito web, si è rivelata particolarmente adatta per un applicativo semplice e minimale il cui unico scopo era la modifica delle immagini da testare.

La parte più consistente dello sviluppo, nonché quella che ha richiesto un maggiore approfondimento di conoscenze, è stata quella riguardante le funzioni per la modifica delle immagini.

Si è usata la libreria *CamanJS*⁴, che contiene diverse funzioni di modifica delle immagini, le quali agiscono trasformando l'immagine in un canvas. Un canvas è un elemento composto dai valori RGB e alpha (trasparenza) di ogni pixel, sul quale è quindi possibile operare modificando un pixel alla volta.

È possibile trovare il codice al link <https://github.com/PinkLaura/pixel-e-percentuali>, e il sito finito su <https://pinklaura.github.io/pixel-e-percentuali/>. È consigliato l'utilizzo del browser *Chrome* per l'accesso al sito, in quanto l'implementazione delle funzioni le rende particolarmente lente da eseguire su altri browser.

2.1 Creazione della pagina web

Nel sito è stato necessario dare la possibilità all'utente di inserire:

- Il **valore** della percentuale che si voleva rappresentare, di seguito **N**;
- Il **colore di sfondo**: bianco o nero;
- La **logica di conversione**: lineare, o S-shaped. Quest'ultima fa sì che i valori centrali rimangano tali mentre rende più estremi i valori in maniera esponenziale, man mano che ci si allontana dal 50%;
- L'**effetto** da applicare all'immagine.

Inoltre, era necessario che fosse possibile inserire N sia tramite valore numerico visualizzandone successivamente il risultato, sia attraverso uno slider che mostrasse le modifiche in tempo reale.

Il sito finale si compone di due semplici pagine: la prima, nella quale è possibile caricare l'immagine alla quale si vuole applicare il filtro, selezionare l'effetto desiderato, il colore di sfondo, la logica di conversione ed il valore di N tramite una textbox numerica.

Dopo aver selezionato gli attributi desiderati è premuto il bottone "Via!", ci si troverà davanti alla nuova immagine, alla quale è stato applicato l'effetto, affiancata dall'immagine originale.

Come è possibile osservare in figura, nella seconda pagina è presente anche uno slider. Esso può essere utilizzato per applicare l'effetto anche alla seconda immagine, consentendo dunque di compararne le varie intensità.

Inserisci gli attributi desiderati

Seleziona l'effetto che desideri testare: Rimozione random di pixel ▾

Carica l'immagine da manipolare:

Scegli file paesaggio_alpi.jpg

Seleziona il colore di sfondo: Bianco ▾

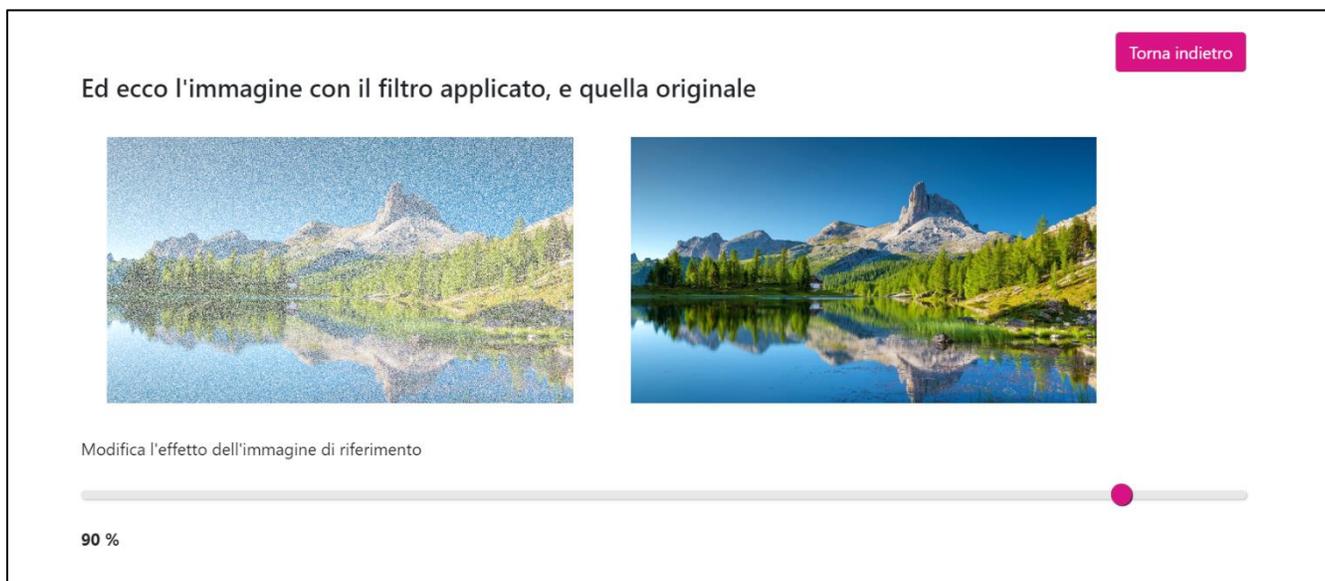
Seleziona la logica di conversione: Lineare ▾

Inserisci la percentuale da rappresentare:

50

Via!

4 Screenshot di un esempio di schermata iniziale



3 Screenshot di un esempio della seconda schermata

2.2 Definizione e sviluppo dei singoli effetti

Il primo passo è stato quello di definire quali soluzioni grafiche avrebbero potuto essere adeguate. È stato necessario individuare degli effetti applicabili ad un'immagine in maniera proporzionale ad un valore scelto (compreso fra 0 e 100), ove i valori estremi fossero chiaramente riconoscibili, e che rappresentassero un giusto equilibrio fra la comunicazione chiara del dato, senza però rimandare direttamente alla sua forma numerica.

Si è dunque scelto di implementare cinque diverse soluzioni grafiche che si è ritenuto potessero rispettare i requisiti.

Le cinque soluzioni erano:

- Rimozione random di pixel
- Rimozione ordinata di pixel
- Scambio casuale di pixel
- Trasparenza
- Blur

Rimozione random di pixel

La funzione di **rimozione casuale dei pixel** è stata ottenuta tramite un ciclo "for" che iterando sui pixel del canvas impone l'*alpha-channel* (valore relativo alla trasparenza, che va da 0. trasparenza totale, a 255, nessuna trasparenza) del pixel a 0 con la probabilità N richiesta.

Era possibile in alternativa "estrarre" casualmente un numero di pixel pari all'N% del totale, controllare che non fossero già stati resi trasparenti, e a quel punto rimuoverli. La generazione di così tante posizioni, e i relativi controlli, risultavano però nettamente più dispendiosi e laboriosi.

Tenuto conto della legge dei grandi numeri, secondo la quale si può essere ragionevolmente sicuri che preso un campione di una certa numerosità una media sperimentale sia sufficientemente vicina a quella teorica, è stato assunto che la percentuale sarebbe stata sufficientemente precisa.

È possibile descrivere l'effetto risultante come una nuvola di pixel bianchi o neri (a seconda del colore di sfondo) che ricopre l'immagine.



6 Esempio di immagine con il 50% di pixel rimossi, e colore di sfondo bianco



5 Esempio di immagine con colore di sfondo nero e solo il 25% dei pixel rimanenti

Rimozione ordinata di pixel

La funzione di **rimozione ordinata dei pixel** è stata ottenuta mediante un ciclo "for" che iterando sui pixel del canvas impone la trasparenza al pixel (*alpha-channel* a 0), su una proporzione di N/10 pixel ogni 10, dove N è la probabilità da rappresentare.

Il numero di pixel rimossi è precisamente pari all' $N\%$, e ne risulta un'immagine con strisce più o meno spesse di pixel mancanti.



8 Esempio di immagine con il 50% dei pixel rimossi e sfondo bianco



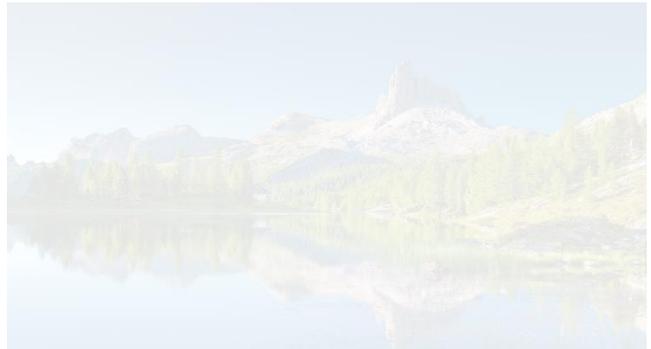
7 Esempio di immagine con solo il 10% dei pixel rimanenti e sfondo nero

Trasparenza

Ogni pixel presente in un canvas presenta quattro attributi: tre di essi rappresentano i valori RGB del pixel, mentre il quarto è il valore di trasparenza (detto *alpha-channel*), misurato su una scala da 0 a 255. La funzione di **trasparenza** assegna all'attributo *alpha* di ogni pixel un valore proporzionale ad N .



9 Esempio di immagine con trasparenza al 90% e sfondo nero



10 Esempio di immagine con trasparenza al 90% e sfondo bianco

Scambio random di pixel

La funzione di **scambio random di pixel** iterando sull'immagine scambia ciascun pixel con uno casuale, con una probabilità di $N/2$. Così facendo otteniamo un'immagine in cui l' $N\%$ dei pixel si trova spostata rispetto alla sua posizione originale.



11 Esempio di immagine con il 25% di pixel scambiati



12 Esempio di immagine con solo il 25% di pixel rimasti al loro posto

Blur

Il **blur**, o *sfocato*, è un effetto che, applicato ad un'immagine, la trasforma come se essa fosse posta al di là di una lente, e fosse dunque uniformemente fuori fuoco. L'effetto classico di blur si effettua colorando ogni pixel con la media pesata, in base alla distanza, di tutti i pixel circostanti.

Non è stato necessario implementare questa funzione, in quanto già presente in molte librerie per l'elaborazione delle immagini. Nello specifico, è stata usata la funzione *stackBlur()* presente nella già citata libreria **CamnJS**.



14 Esempio di Blur al 50%



13 Esempio di immagine rappresentante il 90% mediante un effetto di Blur

3. TEST UTENTE

Dopo aver implementato le soluzioni grafiche, era necessario adottare degli strumenti di valutazione che ci consentissero di capire quanto potessero essere realmente utilizzate in contesti d'uso reali.

Per verificarlo è stata eseguita una **valutazione di usabilità** rispetto alle diverse modalità di visualizzazioni grafiche di percentuali.

L'usabilità è definita dall'ISO (*International Organization for Standardization*) come il grado di efficacia, efficienza e soddisfazione con il quale determinati utenti raggiungono determinati obiettivi in determinati contesti.

Per valutare l'usabilità delle visualizzazioni è stato scelto di costruire un questionario suddiviso in diversi task da portare a termine. I risultati del questionario sono poi stati analizzati statisticamente. Si è reso dunque necessario definire gli **obiettivi** e gli **utenti** che avremmo sottoposto a valutazione.

3.1 Definizione degli obiettivi

L'obiettivo principale delle visualizzazioni grafiche di percentuale è quello di riuscire ad essere **riconoscibili**, pur rimanendo **vaghe**.

La riconoscibilità si riferisce alla capacità di ricondurre la mente dell'utente ad un valore numerico sufficientemente corretto. L'utente deve essere in grado di capire se la percentuale rappresentata è maggiore o minore di una fissata, e di comparare due differenti visualizzazioni.

Tutto ciò non implica che si debba essere in grado di riconoscere l'esatto valore rappresentato, o perderebbe di senso la realizzazione di visualizzazioni vaghe per definizione, è necessario dunque trovare un giusto compromesso.

Nel questionario sono stati creati due diversi tipi di task:

- **Checklist**, presentava un'immagine rappresentante una percentuale ignota, e chiedeva all'utente di dire se riteneva che questa percentuale fosse superiore o inferiore al 50%. L'utente poteva scegliere, oltre che tra "maggiore" e "minore", fra "forse" e "decisamente", in modo da poter esprimere il suo grado di incertezza nella risposta.
- Il secondo task era composto di un semplice **Slider**, con il quale all'utente era chiesto di indicare con precisione quale fosse il valore della percentuale rappresentata. Se con il task Checklist si analizza l'intuitività della rappresentazione, in questo caso l'obiettivo del task è comprendere se questo tipo di rappresentazioni riesca ad essere sufficientemente preciso.

Secondo te a quanto corrisponde la "purezza" della immagine precedente in una scala tra 0 e 100%?

Muovi lo slider

58

0 100

? (muovi lo slider per indicare il valore che si avvicina di più alla purezza reale)

Osserva questa immagine.



Secondo te, rispetto alla purezza del 50%, questa immagine ha purezza:

- Decisamente Maggiore
- Forse Maggiore
- Forse Minore
- Decisamente Minore
- Non saprei

? Ricorda, più l'effetto è marcato e minore è la purezza della immagine.

16 Esempio di task di tipo Checklist

È stata inoltre selezionata una serie di diverse percentuali da indovinare, in modo da testare la comprensione sia di valori estremi sia di valori più centrali.

Inoltre, sono stati preparati tre diversi questionari, così da diversificare le associazioni percentuale-effetto, e ridurre al minimo eventuali bias.

Per rendere inoltre più scorrevole il questionario, e non appesantirlo troppo, si è scelto di ridurre a il gruppo di effetti da analizzare ad un numero di tre. Gli effetti ritenuti più adatti, in quanto molto diversi fra loro, sono stati la rimozione random di pixel (d'ora in avanti, **rumore**), la **trasparenza** su sfondo nero, e il **blur**.

I valori percentuali sono risultati quindi distribuiti nei diversi questionari 1, 2 e 3 come riportato in tabella.

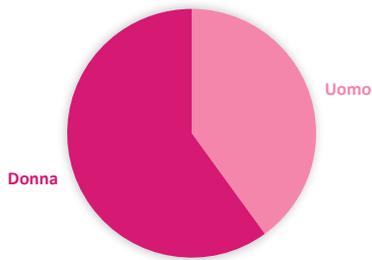
%	TRASPARENZA	BLUR	RUMORE
10	3	1	1
25	1	3	2
40	2	1	3
60	3	2	2
75	2	3	1
90	1	2	3

3.2 Definizione degli utenti

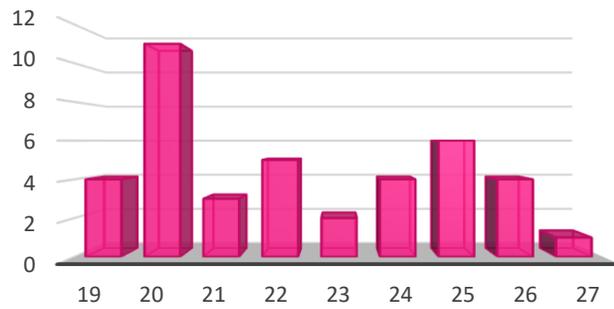
Il campione ideale di utenti sarebbe stato composto interamente da medici. Essendo però difficile reperire una quarantina di medici disponibili, è stato scelto di sottoporre al test **40 studenti di medicina**.

In questo modo si è cercato di considerare un campione abbastanza simile a quello ideale, ma di più facile approccio.

GENERE DEI RISPONDENTI



ETÀ DEI RISPONDENTI



Non abbiamo reputato rilevante una suddivisione del campione per particolari fasce d'età, in quanto si tratta interamente di giovani di età compresa fra i 19 e i 27 anni, e una differenza di età di qualche anno non influisce sullo svolgimento dei task.

4. RISULTATI TEST UTENTE

Durante lo svolgimento dei task è stata data la possibilità agli utenti di domandare aiuto in caso di difficoltà, ciò non è mai accaduto. Tutto lo svolgimento è stato ripreso, sia attraverso la webcam sia dal punto di vista di ciò che accade sullo schermo, in modo da poter notare eventuali particolari difficoltà degli utenti, e misurare i loro tempi di esecuzione.

Le misurazioni relative ai task sono state analizzate mediante dei test statistici, che hanno studiato due differenti dimensioni di usabilità della modalità di visualizzazione:

- Efficacia
- Efficienza

L'**efficacia** può essere definita come la misura di accuratezza e completezza con cui gli utenti possono raggiungere i loro obiettivi grazie al sistema. Nel nostro specifico caso, considereremo l'efficacia tramite il numero di **successi** e **fallimenti** nel completamento dei task.

L'**efficienza** è misura invece le risorse spese in relazione all'accuratezza e completezza degli obiettivi raggiunti. La risorsa che interessa a noi è il tempo, e per valutare l'efficacia del nostro modello utilizzeremo dei test statistici in grado di confrontare i **tempi** di esecuzione.

È stato scelto, per efficacia ed efficienza, di valutarne sia le misure assolute che quelle relative. Questo per capire in primo luogo se l'intero sistema di visualizzazioni alternative di percentuali, sia usabile. In secondo luogo, per comprendere se ci fossero differenze statisticamente significative nell'usabilità delle visualizzazioni ottenute applicando i tre diversi effetti.

In tutti i test il livello di significatività α è fissato al 5%. Dunque, rigetteremo l'ipotesi nulla se il p-value sarà inferiore a questa soglia.

4.1 Checklist task: efficacia

È riportata di seguito la matrice degli errori del **task checklist**. Nel task era presente la risposta "forse maggiore/minore", che avrebbe potuto essere interpretata come un *mezzo successo*. Parlando con gli utenti, però, ci si è resi conto che se alcuni avevano interpretato la risposta come un modo per esprimere la propria incertezza, altri l'avevano intesa come una scala ordinata, dove la risposta "forse" indicava solo una certa vicinanza al 50%, nonostante gli utenti fossero certi di dove posizionarla.

Dunque, è stato ritenuto opportuno indicare un task come *sbagliato* solo nel caso in cui si fosse effettivamente sbagliata la risposta, e altrimenti considerarlo giusto.

Nella matrice sono presenti 80 risposte per ogni task, questo è dovuto al fatto che i 40 utenti hanno dovuto riconoscere due diverse percentuali in ogni task, e sono state analizzate tutte le risposte complessivamente.

CHECKLIST TASK						
ID	BLUR		RUMORE		TRASPARENZA	
7						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
93						
95						

96						
97						
98						
99						
100						
101						
102						
103						
104						
105						
106						
107						
108						

Efficacia assoluta: l'Error rate

L'**Error Rate**, o tasso di errore, è il rapporto fra i task falliti e quelli completati con successo. Dovrebbe assestarsi sotto l'1% in caso di alta familiarità con l'applicativo, e mantenersi comunque sotto al 5% se la familiarità è bassa.

L'Error rate del task Checklist è dell'8%.

	Successi	Insuccessi	Error rate
BLUR	70	10	0,13
RUMORE	76	4	0,05
TRASPARENZA	75	5	0,06
	221	19	0,08

È stato deciso di valutare sia l'efficacia della rappresentazione di percentuale realizzata con ciascun diverso effetto, sia quella dell'intero sistema.

Per misurare l'efficacia assoluta ci si è avvalsi di un test del chi quadro, con il quale è possibile determinare se la differenza osservata tra l'Error Rate calcolato e il 5% sia reale o dovuta al caso. L'ipotesi nulla, dunque, è che non ci sia differenza statisticamente significativa tra l'Error Rate e il 5% nell'eseguire il task Checklist.

Blur

Un **test del Chi Quadro** è stato eseguito per confrontare tra loro una percentuale accettabile del 5% e la percentuale di insuccessi (10/80) degli utenti nel compiere il **task checklist** usando il filtro *Blur*.

In base a tale test si è trovata una **DIFFERENZA STATISTICAMENTE NON SIGNIFICATIVA** tra il tasso di errore e il 5%; $p = 0.0942$.

Questo risultato suggerisce che il numero di errori non sia significativamente maggiore del 5%, e dunque che il task Checklist sia sufficientemente efficace relativamente all'effetto Blur.

Rumore

Un **test del Chi Quadro** è stato eseguito per confrontare tra loro una percentuale accettabile del 5% e la percentuale di insuccessi (4/80) degli utenti nel compiere il **task checklist** usando il filtro *Rumore*.

In base a tale test **NON SI È TROVATA DIFFERENZA** tra il tasso di errore e il 5%; $p = 1.0000$.

Questo risultato suggerisce che il numero di errori non sia significativamente maggiore del 5%, e dunque che il task Checklist sia sufficientemente efficace relativamente all'effetto Rumore.

Trasparenza

Un **test del Chi Quadro** è stato eseguito per confrontare tra loro una percentuale accettabile del 5% e la percentuale di insuccessi (5/80) degli utenti nel compiere il **task checklist** usando il filtro *Trasparenza*.

In base a tale test si è trovata una **DIFFERENZA STATISTICAMENTE NON SIGNIFICATIVA** tra il tasso di errore e il 5%; $p = 0.7323$.

Questo risultato suggerisce che il numero di errori non sia significativamente maggiore del 5%, e dunque che il task Checklist sia sufficientemente efficace relativamente all'effetto Trasparenza.

Intero sistema

Un **test del Chi Quadro** è stato eseguito per confrontare tra loro una percentuale accettabile del 5% e la percentuale di insuccessi (19/240) degli utenti nel compiere il **task checklist** usando il sistema.

In base a tale test si è trovata una **DIFFERENZA STATISTICAMENTE NON SIGNIFICATIVA** tra il tasso di errore e il 5%; $p = 0.1940$.

Questo risultato suggerisce che il numero di errori non sia significativamente maggiore del 5%, e dunque che il task Checklist sia sufficientemente efficace relativamente all'intero sistema.

Efficacia relativa

Per valutare l'efficacia relativa, abbiamo confrontato fra loro il tasso di errori dei tre differenti metodi di visualizzazione delle percentuali, per vedere se essi sono significativamente diversi. **L'ipotesi nulla, dunque, è che non ci sia differenza statisticamente significativa tra il numero di successi nell'eseguire il task Checklist ed insuccessi ottenuti utilizzando visualizzazioni con effetti differenti.**

Un **TEST DEL CHI QUADRO** è stato eseguito per confrontare il rapporto fra successi ed errori nel compiere il **task checklist** usando i tre filtri.

In base a tale test, si è trovata una **DIFFERENZA STATISTICAMENTE NON SIGNIFICATIVA** riguardo al numero di errori, tra *Blur*, *Rumore* e *Trasparenza*; $p = 0.170018$.



task Checklist a seconda del tipo di effetto.

Un **TEST DEL CHI QUADRO** è stato eseguito anche per confrontare il rapporto fra successi ed errori nel compiere il task **checklist** usando solo i filtri *Blur* e *Rumore*.

In base a tale test, si è trovata una **DIFFERENZA STATISTICAMENTE NON SIGNIFICATIVA** riguardo al numero di errori, tra *Blur* e *Rumore*; $p = 0.093212$.

Questi risultati suggeriscono che gli utenti non commettono un numero di errori significativamente diverso nell'esecuzione del

4.2 Slider task: efficacia

L'esecuzione del **task slider** richiedeva di selezionare un valore tra 1 e 100, che corrispondesse al valore della percentuale rappresentata dall'immagine.

Si è ritenuto che giudicare corretto un task solo nel caso in cui la percentuale fosse stata azzeccata perfettamente fosse eccessivamente limitante, dunque è stato stabilito un intervallo di valori all'interno del quale il valore inserito sarebbe stato giudicato corretto, altrimenti il task sarebbe stato valutato come fallimentare.

L'intervallo scelto ha come centro il valore esatto, e ampiezza determinata dalla deviazione standard delle risposte. Di seguito è riportata la matrice successi/insuccessi relativa al task.

SLIDER TASK						
ID	BLUR		RUMORE		TRASPARENZA	
7						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
93						
95						
96						
97						

98						
99						
100						
101						
102						
103						
104						
105						
106						
107						
108						

Efficacia assoluta: l'Error rate

L'Error rate del task Slider è del 44%.

	Successi	Insuccessi	Error rate
BLUR	45	35	0,44
RUMORE	48	32	0,40
TRASPARENZA	41	39	0,49
	134	106	0,44

È stato deciso di valutare sia l'efficacia della rappresentazione di percentuale realizzata con ciascun diverso effetto, sia quella dell'intero sistema.

Per misurare l'efficacia assoluta ci si è avvalsi di un test del chi quadro, con il quale è possibile determinare se la differenza osservata tra l'Error Rate calcolato e il 5% sia reale o dovuta al caso. L'ipotesi nulla, dunque, è che non ci sia differenza statisticamente significativa tra l'Error Rate e il 5% nell'eseguire il task Slider.

Blur

Un test del Chi Quadro è stato eseguito per confrontare tra loro una percentuale accettabile del 5% e la percentuale di insuccessi (35/80) degli utenti nel compiere il task Slider usando il filtro *Blur*.

In base a tale test si è trovata una **DIFFERENZA STATISTICAMENTE SIGNIFICATIVA** tra il tasso di errore e il 5%; $p < 0.0001$.

Questo risultato suggerisce che il numero di errori sia significativamente maggiore del 5%, e dunque che il task Slider non sia sufficientemente efficace relativamente all'effetto *Blur*.

Rumore

Un **test del Chi Quadro** è stato eseguito per confrontare tra loro una percentuale accettabile del 5% e la percentuale di insuccessi (32/80) degli utenti nel compiere il **task Slider** usando il filtro *Rumore*.

In base a tale test si è trovata una **DIFFERENZA STATISTICAMENTE SIGNIFICATIVA** tra il tasso di errore e il 5%; $p < 0.0001$.

Questo risultato suggerisce che il numero di errori sia significativamente maggiore del 5%, e dunque che il task Slider non sia sufficientemente efficace relativamente all'effetto Rumore.

Trasparenza

Un **test del Chi Quadro** è stato eseguito per confrontare tra loro una percentuale accettabile del 5% e la percentuale di insuccessi (39/80) degli utenti nel compiere il **task Slider** usando il filtro *Trasparenza*.

In base a tale test si è trovata una **DIFFERENZA STATISTICAMENTE SIGNIFICATIVA** tra il tasso di errore e il 5%; $p < 0.0001$.

Questo risultato suggerisce che il numero di errori sia significativamente maggiore del 5%, e dunque che il task Slider non sia sufficientemente efficace relativamente all'effetto Trasparenza.

Intero sistema

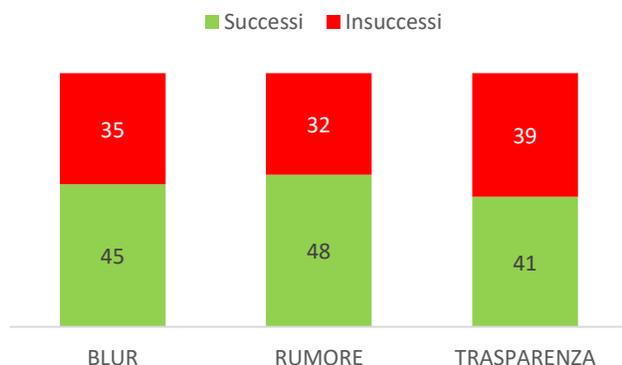
Un **test del Chi Quadro** è stato eseguito per confrontare tra loro una percentuale accettabile del 5% e la percentuale di insuccessi (106/240) degli utenti nel compiere il **task Slider** usando il sistema.

In base a tale test si è trovata una **DIFFERENZA STATISTICAMENTE SIGNIFICATIVA** tra il tasso di errore e il 5%; $p < 0.0001$.

Questo risultato suggerisce che il numero di errori sia significativamente maggiore del 5%, e dunque che il task Slider non sia sufficientemente efficace relativamente all'intero sistema.

Efficacia relativa

L'ipotesi nulla considerata è che non ci sia differenza statisticamente significativa tra il numero di successi nell'eseguire il task Slider ed insuccessi ottenuti utilizzando visualizzazioni con effetti differenti.



Un **TEST DEL CHI QUADRO** è stato eseguito per confrontare il rapporto fra successi ed errori nel compiere il task **slider** usando i tre filtri.

In base a tale test, si è trovata una **DIFFERENZA STATISTICAMENTE NON SIGNIFICATIVA** riguardo al numero di errori, tra *Blur*, *Rumore* e *Trasparenza*; $p = 0.535167$.

Un **TEST DEL CHI QUADRO** è stato eseguito anche per confrontare il rapporto fra successi ed errori nel compiere il task **slider** usando i filtri *Rumore* e *Trasparenza*.

In base a tale test, si è trovata una **DIFFERENZA STATISTICAMENTE NON SIGNIFICATIVA** riguardo al numero di errori tra *Rumore* e *Trasparenza*; $p = 0.265336$.

Questi risultati suggeriscono che gli utenti non commettono un numero di errori significativamente diverso nell'esecuzione del task Slider a seconda del tipo di effetto.

4.3 Efficienza

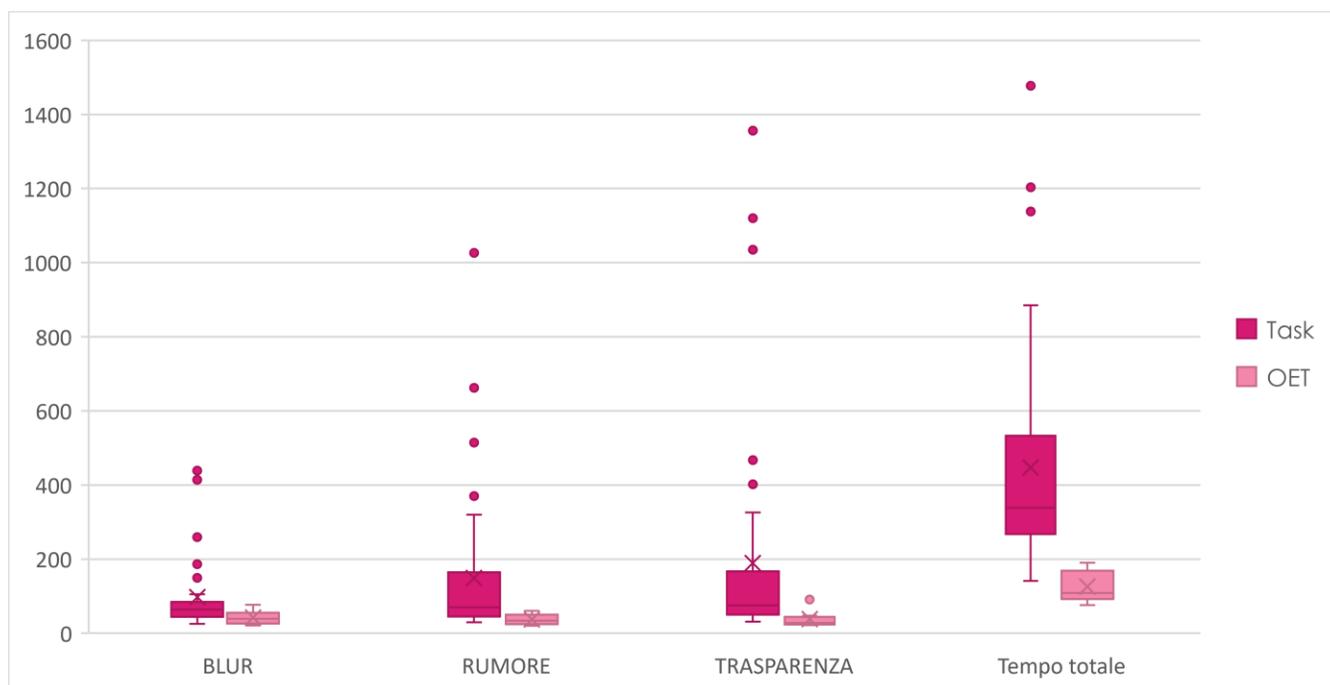
Per misurare l'efficienza abbiamo confrontato tra loro i **tempi di esecuzione** dei task. Abbiamo scelto di misurare sia l'efficienza assoluta di questa modalità di visualizzazione, sia l'efficienza relativa.

Efficienza assoluta: l'OET

Per misurare l'efficienza assoluta è stato calcolato l'**Optimal Execution Time**. Si tratta del tempo necessario allo svolgimento dei task per utenti che conoscono già perfettamente questa modalità di visualizzazione delle percentuali. È stato ottenuto registrando i tempi impiegati dagli stessi valutatori, che hanno ripetuto più volte l'esecuzione dei task a diverse velocità.

Se non si osserva una differenza significativa tra l'OET e i tempi campionari, si può definire il sistema efficiente.

L'ipotesi nulla, dunque, è che **non ci sia differenza statisticamente significativa tra i tempi di esecuzione dei gruppi di task relativi ad ogni effetto e l'OET**.



Blur

Un T-TEST per **campioni indipendenti** è stato eseguito per confrontare il tempo di esecuzione del campione con l'OET nello svolgere il gruppo di domande relative all'effetto Blur.

In base a tale test, si è trovata una **DIFFERENZA STATISTICAMENTE NON SIGNIFICATIVA** fra i tempi di esecuzione dei task del campione ($97,17 \pm 32,81$) e l'OET ($42,08 \pm 14,13$), relativi all'effetto *Blur*; $t = 1.54$; $p = 0.129312$.

Questo risultato suggerisce che una bassa conoscenza del sistema non rallenti eccessivamente i nuovi utenti nella comprensione delle percentuali rappresentate con l'effetto Blur.

Rumore

Un T-TEST per **campioni indipendenti** è stato eseguito per confrontare il tempo di esecuzione del campione con l'OET nello svolgere il gruppo di domande relative all'effetto Rumore.

In base a tale test, si è trovata una **DIFFERENZA STATISTICAMENTE NON SIGNIFICATIVA** fra i tempi di esecuzione dei task del campione ($148,77 \pm 61,27$) e l'OET ($37,32 \pm 11,13$), relativi all'effetto *Rumore*; $t = 1,68$; $p = 0,100282$.

Questo risultato suggerisce che una bassa conoscenza del sistema non rallenti eccessivamente i nuovi utenti nella comprensione delle percentuali rappresentante con l'effetto Rumore.

Trasparenza

Un T-TEST per **campioni indipendenti** è stato eseguito per confrontare il tempo di esecuzione del campione con l'OET nello svolgere il gruppo di domande relative all'effetto Trasparenza.

In base a tale test, si è trovata una **DIFFERENZA STATISTICAMENTE NON SIGNIFICATIVA** fra i tempi di esecuzione dei task del campione ($189,73 \pm 93,30$) e l'OET ($37,77 \pm 16,81$), relativi all'effetto *Trasparenza*; $t = 1,50$; $p = 0,140009$.

Questo risultato suggerisce che una bassa conoscenza del sistema non rallenti eccessivamente i nuovi utenti nella comprensione delle percentuali rappresentante con l'effetto Trasparenza.

Tempo totale

Un T-TEST per **campioni indipendenti** è stato eseguito per confrontare il tempo di esecuzione del campione con l'OET nello svolgere l'intero gruppo di task.

In base a tale test, si è trovata una **DIFFERENZA STATISTICAMENTE SIGNIFICATIVA** fra i tempi di esecuzione dei task del campione ($447,13 \pm 96,34$) e l'OET ($125,86 \pm 31,64$), $t = 3,07$; $p = 0,003553$.

Questo risultato suggerisce che una bassa conoscenza del sistema rallenti eccessivamente i nuovi utenti nella comprensione dei metodi alternativi di visualizzazione di percentuali, di conseguenza possiamo ritenere che la modalità di visualizzazione non sia sufficientemente efficiente.

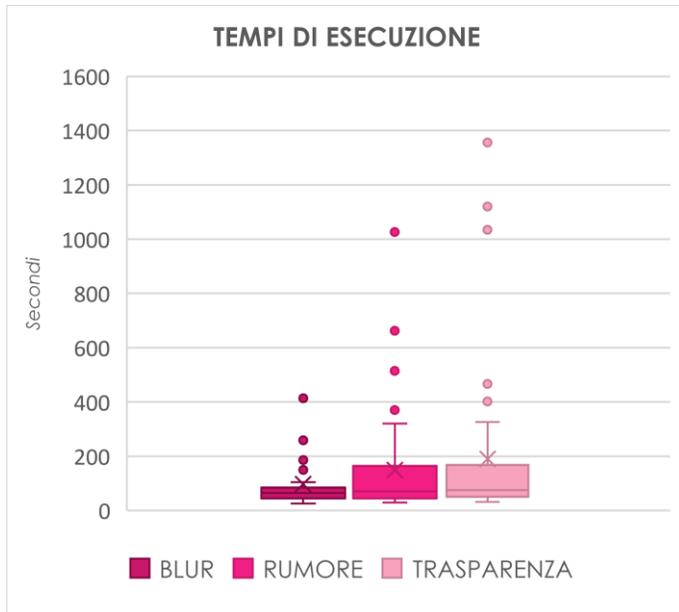
Nell'ultimo caso, l'**ipotesi nulla presa in considerazione era che non ci fosse una differenza significativa fra il tempo totale impiegato dal campione, e l'OET.**

È interessante notare come non si siano rilevate differenze fra i tre effetti presi singolarmente, ma siano invece state trovate nel confronto dei tempi totali.

Efficienza relativa

Oltre all'efficienza assoluta del sistema, abbiamo voluto provare a vedere se ci fossero differenze tra i tempi di esecuzione dei vari gruppi di domande, ossia la coppia di task relativa ad ogni effetto.

L'ipotesi nulla considerata è che non ci sia differenza statisticamente significativa tra tempi misurati nell'esecuzione dei tre differenti gruppi di task.



Un TEST ANOVA per campioni dipendenti è stato eseguito per confrontare i tempi di esecuzione nel compiere i due task usando i tre filtri.

In base a tale test, si è trovata una **DIFFERENZA STATISTICAMENTE NON SIGNIFICATIVA** fra i tempi di esecuzione di *Blur* ($97,17 \pm 32,81$), *Rumore* ($148,77 \pm 61,27$) e *Trasparenza* ($189,73 \pm 93,30$); $p = 0.055051$.

Questo risultato suggerisce che gli utenti non impiegano un numero significativamente diverso di secondi nell'esecuzione dei task a seconda del tipo di effetto.

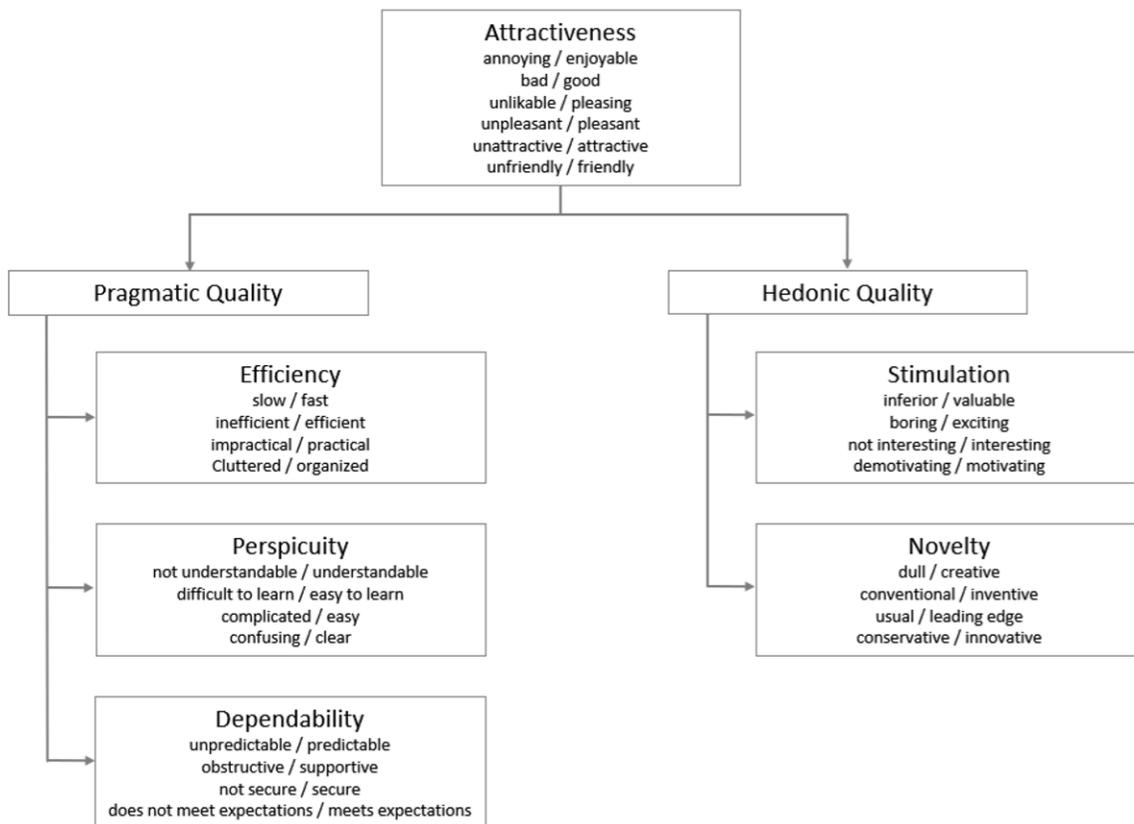
5. QUESTIONARIO DI VALUTAZIONE

Al termine dei tasks è stato somministrato ad ogni utente un questionario, per poter raccogliere dati relativi alla loro esperienza utente, con l'obbiettivo di valutarne la **soddisfazione**.

Il modello scelto è stato quello dell'**User Experience Questionnaire**. È stato scelto questo questionario per la sua chiarezza e per la sua brevità.

Avendo somministrato il questionario a degli sconosciuti dopo dei task utente che richiedevano un notevole livello di concentrazione, somministrare un questionario dalla durata superiore ai 10 minuti avrebbe significato doversi scontrare con il "fatigue bias". Si tratta di un bias dovuto alla tendenza di alcuni utenti a rispondere in maniera randomica a causa della stanchezza, creando così rumore nelle risposte. Un altro rischio dovuto alla stanchezza è il "bias di tendenza centrale", ossia la tendenza dei rispondenti a dare punteggi neutri, meno faticosi da assegnare.

Il questionario misura 6 diverse scale, così strutturate:



Inoltre, per meglio adattarlo al nostro progetto, abbiamo apportato le seguenti modifiche:

1. È stato eliminato il quesito "di grande valore - di poco valore", ritenuto superfluo e poco chiaro
2. È stato ridotto l'intervallo delle risposte da 1 - 7 a 1 - 5, per limitare la dispersione e agevolare l'utente nelle risposte
3. È stato aggiunto un quesito, per misurare il Net Promoter Score

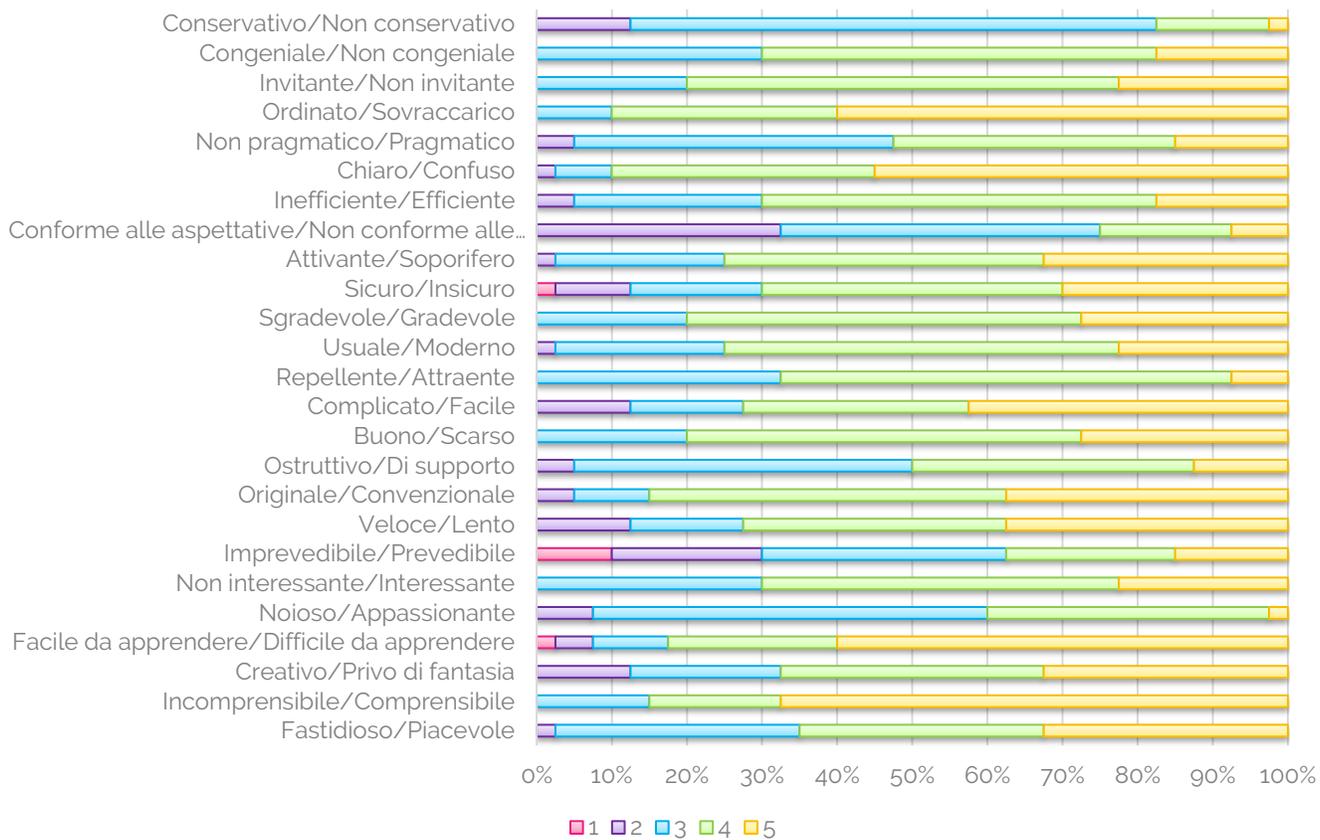
4. È stata aggiunta una domanda aperta, per raccogliere eventuali suggerimenti

È possibile consultare il questionario completo nell'Appendice.

5.1 Distribuzione Risposte

In questa sezione mostreremo le distribuzioni delle risposte per ogni quesito. Analizzeremo tali risposte nel paragrafo successivo, suddividendole per scala.

Nr	Domanda	1	2	3	4	5	Scala
1	Fastidioso/Piacevole	0	1	13	13	13	Attrattività
2	Incomprensibile/Comprensibile	0	0	6	7	27	Apprendibilità
3	Creativo/Privo di fantasia	0	5	8	14	13	Originalità
4	Facile da apprendere/Difficile da apprendere	1	2	4	9	24	Apprendibilità
5	Noioso/Appassionante	0	3	21	15	1	Stimolazione
6	Non interessante/Interessante	0	0	12	19	9	Stimolazione
7	Imprevedibile/Prevedibile	4	8	13	9	6	Controllabilità
8	Veloce/Lento	0	5	6	14	15	Efficienza
9	Originale/Convenzionale	0	2	4	19	15	Originalità
10	Ostruttivo/Di supporto	0	2	18	15	5	Controllabilità
11	Buono/Scarso	0	0	8	21	11	Attrattività
12	Complicato/Facile	0	5	6	12	17	Apprendibilità
13	Repellente/Attrattante	0	0	13	24	3	Attrattività
14	Usuale/Moderno	0	1	9	21	9	Originalità
15	Sgradevole/Gradevole	0	0	8	21	11	Attrattività
16	Sicuro/Insicuro	1	4	7	16	12	Controllabilità
17	Attivante/Soporifero	0	1	9	17	13	Stimolazione
18	Conforme alle aspettative/Non conforme alle aspettative	0	13	17	7	3	Controllabilità
19	Inefficiente/Efficiente	0	2	10	21	7	Efficienza
20	Chiaro/Confuso	0	1	3	14	22	Apprendibilità
21	Non pragmatico/Pragmatico	0	2	17	15	6	Efficienza
22	Ordinato/Sovraccarico	0	0	4	12	24	Efficienza
23	Invitante/Non invitante	0	0	8	23	9	Attrattività
24	Congeniale/Non congeniale	0	0	12	21	7	Attrattività
25	Conservativo/Non conservativo	0	5	28	6	1	Originalità



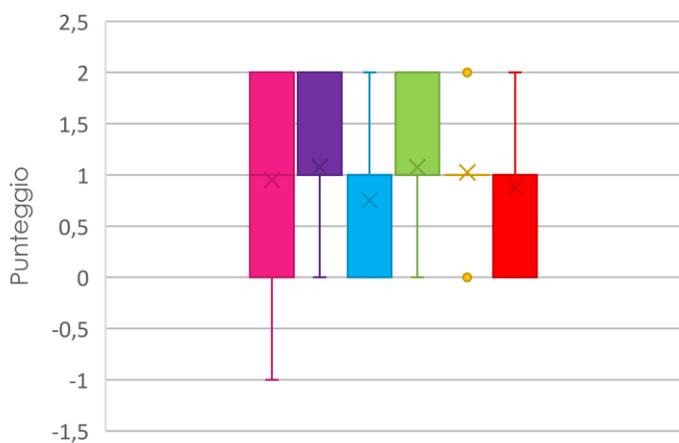
5.2 Analisi Per scala

I risultati ottenuti sui quesiti sono stati analizzati divisi per scala di appartenenza.

Nelle seguenti pagine verranno riportati i boxplots relativi alle singole scale. Rappresentano il valore minimo, primo quartile, mediana, terzo quartile, punteggio massimo ed eventuali outliers relativi ad ogni domanda. Inoltre verranno riportate media e varianza relative alla singola scala.

Le risposte, inizialmente in scala da 1 a 5, sono state convertite in una scala che va da -2 a 2, per permettere un'interpretazione speculare dei dati.

Attrattività



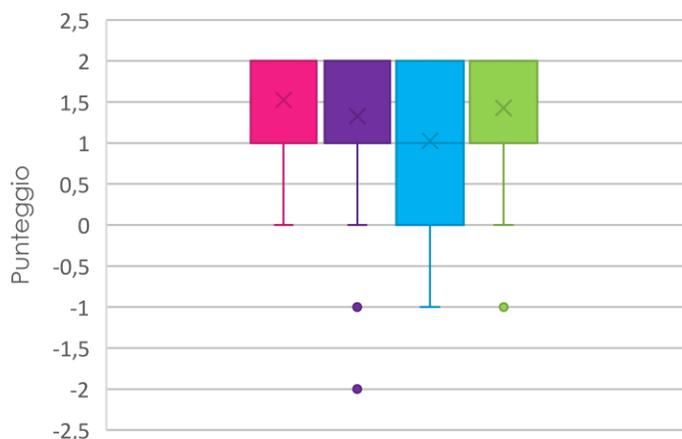
Attrattività	
Media	Varianza
0,96	0,25

Quesito	Minimo	Primo quartile	Mediana	Terzo quartile	Massimo	Media	Varianza
Fastidioso - Piacevole	-1	0	1	2	2	0,95	0,77
Buono - Scarso	0	1	1	2	2	1,08	0,48
Repellente - Attraente	0	0	1	1	2	0,75	0,35
Sgradevole - Gradevole	0	1	1	2	2	1,08	0,48
Invitante - Non invitante	0	1	1	1	2	1,03	0,44
Congeniale - Non congeniale	0	0	1	1	2	0,88	0,47

Nell'ambito dell'attrattività i quesiti che hanno avuto punteggi migliori sono stati "Buono - Scarso" e "Sgradevole - Gradevole", entrambi con media 1,08 e varianza 0,48.

Il punteggio peggiore (ma comunque positivo) è stato attribuito al quesito "Repellente - Attraente", con media 0,75 e varianza 0,35.

Apprendibilità

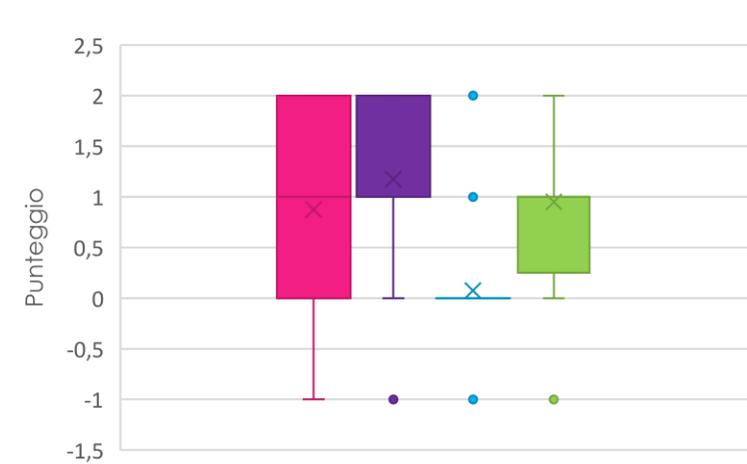


Attrattività	
Media	Varianza
1,33	0,41

Quesito	Minimo	Primo quartile	Mediana	Terzo quartile	Massimo	Media	Varianza
Incomprensibile - Comprensibile	0	1	2	2	2	1,53	0,56
Facile da apprendere - Difficile da apprendere	-2	1	2	2	2	1,33	1,05
Complicato - Facile	-1	0	1	2	2	1,03	1,10
Chiaro - Confuso	-1	1	2	2	2	1,43	0,56

Nell'ambito dell'apprendibilità il quesito che ha avuto più successo è stato quello sulla comprensibilità ("Incomprensibile - Comprensibile") con un punteggio medio di 1,53 e una varianza di 0,56. Mentre il punteggio peggiore è stato registrato dal quesito "Chiaro - Confuso", con una media di 1,43 e una varianza di 0,56. Anche in questo caso, il risultato peggiore è comunque positivo.

Originalità



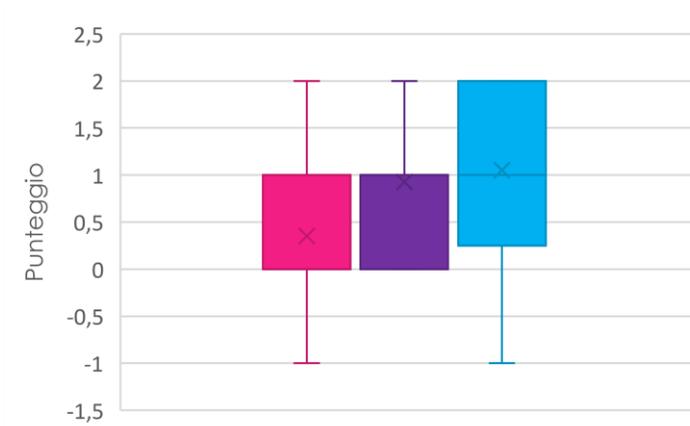
Attrattività	
Media	Varianza
0,77	0,29

Quesito	Minimo	Primo quartile	Mediana	Terzo quartile	Massimo	Media	Varianza
Creativo - Privo di fantasia	-1	0	1	2	2	0,88	1,04
Originale - Convenzionale	-1	1	1	2	2	1,18	0,66
Usuale - Moderno	-1	0,75	1	1	2	0,95	0,56
Conservativo - Non conservativo	-1	0	0	0	2	0,08	0,38

La scala dell'originalità ha avuto il suo miglior punteggio con il quesito "Originale – Convenzionale", con un punteggio medio di 1,18 e varianza di 0,66. Il risultato peggiore è stato attribuito al quesito "Conservativo – Non conservativo", con una media di 0,08 e varianza di 0,38.

Su quest'ultimo ha influito il bias di tendenza centrale. Difatti, salvo alcuni outliers, il valore del primo quartile, mediana e terzo quartile è stato di 0, come è evidente dal boxplot.

Stimolazione

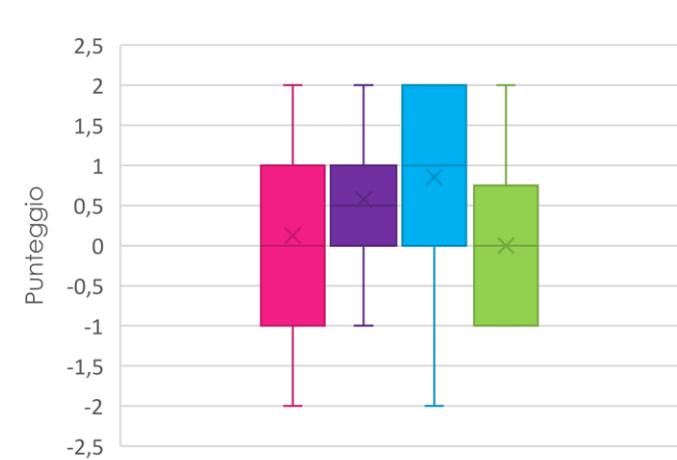


Attrattività	
Media	Varianza
0,96	0,25

Quesito	Minimo	Primo quartile	Mediana	Terzo quartile	Massimo	Media	Varianza
Noioso - Appassionante	-1	0	0	1	2	0,35	0,44
Non interessante - Interessante	0	0	1	1	2	0,93	0,53
Attivante - Soporifero	-1	0,75	1	2	2	1,05	0,66

La scala della stimolazione ha avuto il suo miglior punteggio nel quesito "Attivante – Soporifero", con una media di 1,05 e varianza di 0,66. Quello peggiore è stato registrato dal quesito "Noioso – Appassionante", con una media di 0,35 e varianza di 0,44. In quest'ultimo ha agito il bias di tendenza centrale, infatti 21 rispondenti su 40 hanno dato una risposta neutra.

Controllabilità



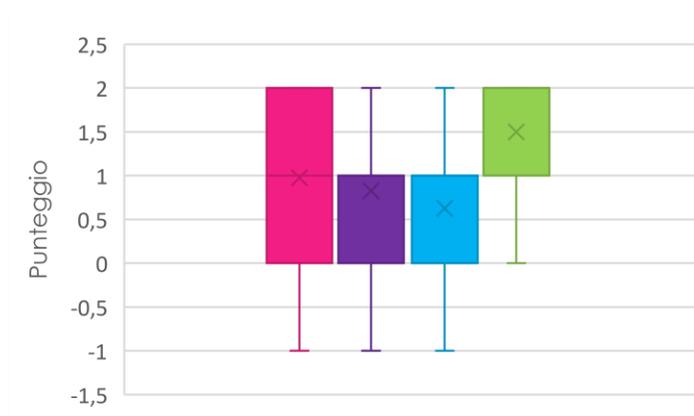
Attrattività	
Media	Varianza
0,39	0,32

Quesito	Minimo	Primo quartile	Mediana	Terzo quartile	Massimo	Media	Varianza
Imprevedibile - Prevedibile	-2	-1	0	1	2	0,13	1,45
Ostruttivo - Di supporto	-1	0	0,5	1	2	0,58	0,61
Sicuro - Insicuro	-2	0	1	2	2	0,85	1,11
Conforme alle aspettative - Non conforme alle aspettative	-1	-1	0	0,25	2	0,00	0,82

Nell'ambito della controllabilità il punteggio migliore è stato dato al quesito "Sicuro – Insicuro", con una media di 0,85 e varianza di 1,11. Quello peggiore dal quesito "Conforme alle aspettative - Non conforme alle aspettative", con media 0,00 e varianza di 0,82.

Come si intuisce anche dall'elevata varianza, quest'ultimo quesito, pur avendo media pari a 0, non è stato inficiato dal bias di tendenza centrale per la maggioranza dei rispondenti, nonostante una parte significativa (17 su 40) abbia dato una risposta neutra.

Efficienza



Attrattività	
Media	Varianza
0,98	0.22

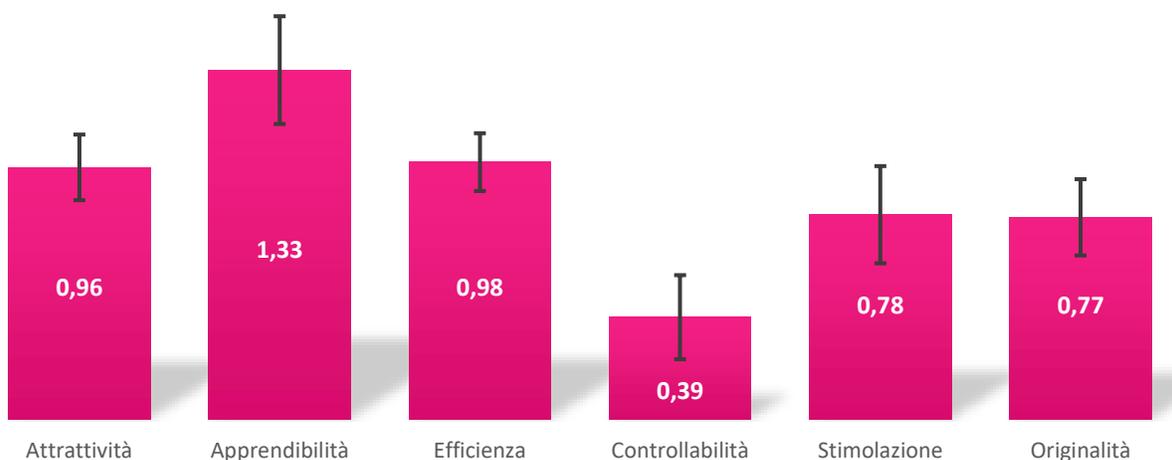
Quesito	Minimo	Primo quartile	Mediana	Terzo quartile	Massimo	Media	Varianza
Veloce - Lento	-1	0	1	2	2	0,98	1,05
Inefficiente - Efficiente	-1	0	1	1	2	0,83	0,61
Non pragmatico - Pragmatico	-1	0	1	1	2	0,63	0,65
Ordinato - Sovraccarico	0	1	2	2	2	1,50	0,46

Nell'ambito dell'efficienza il miglior punteggio è stato registrato dal quesito "Ordinato – Sovraccarico", con una media di 1,50 e varianza di 0,46. Quello peggiore dal quesito "Non pragmatico – Pragmatico", con media pari a 0,63 e varianza 0,65.

5.3 Confronto fra scale

Di seguito i risultati del confronto tra i dati relativi alle 6 diverse scale.

SCALA	MEDIA	VARIANZA
Attrattività	0,96	0,25
Apprendibilità	1,33	0,41
Efficienza	0,98	0,22
Controllabilità	0,39	0,32
Stimolazione	0,78	0,37
Originalità	0,77	0,29



Si può notare che la media è stata calcolata sulle medie delle risposte dei singoli rispondenti, divise per scala. Quindi la varianza si riferisce alla varianza delle medie delle risposte raggruppate per scala. Pertanto la significatività statistica di questa comparazione è minore rispetto a quella dell'analisi domanda per domanda fatta in precedenza.

In generale, il valore della varianza è abbastanza basso, questo dovuto al fatto che le risposte tendono ad essere concentrate attorno al valore medio, mentre si trovano pochi valori estremi.

5.4 Net Promoter score

Il **Net Promoter Score** (NPS) è un indicatore usato per misurare la fedeltà o l'apprezzamento di un utente nei confronti di un'azienda o di un applicativo. Esso misura la proporzione di **"promotori"** di un prodotto o un'azienda, rispetto al numero di **"detrattori"**.

Il NPS si basa su un'unica domanda:

"Con quale probabilità consiglieresti questo prodotto/servizio/sito a un amico o a un collega?"

Modificata per questa valutazione in:

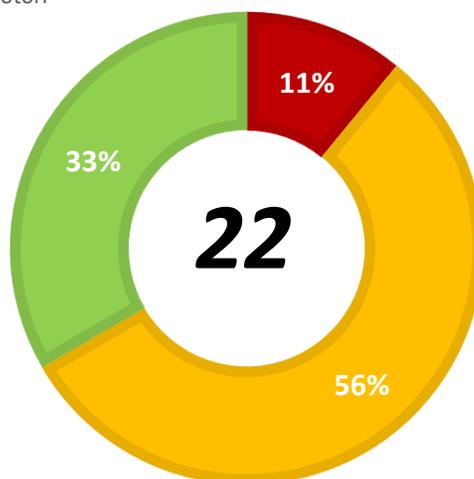
"Con quale probabilità consiglieresti l'utilizzo di questo tipo di immagini per visualizzare delle percentuali?"

Le risposte fornite dagli utenti sono classificate come segue:

- **0-6 = Detrattori:** utenti insoddisfatti che potrebbero danneggiare l'azienda/applicazione attraverso un passaparola negativo;
- **7-8 = Passivi:** utenti soddisfatti ma indifferenti, non promuoveranno, ma nemmeno detrarranno l'azienda/applicazione. Potrebbero essere influenzati da eventuali concorrenti;
- **9-10 = Promotori:** sono gli utenti che sono pienamente soddisfatti e consiglierebbero sicuramente l'applicazione ad altre persone.

Il Net Promoter Score viene calcolato sottraendo la percentuale di detrattori alla percentuale di promotori ottenuta. Il risultato non viene però espresso in punti percentuale, ma come numero assoluto compreso tra -100 e +100.

■ Detrattori ■ Passivi ■ Promotori



Il Net Promoter Score è risultato essere 22, indicativo del fatto che i promotori di questo metodo di visualizzazione sono di più rispetto ai detrattori. Ma la maggioranza degli utenti rimane passiva.

6. Conclusioni

Per quanto riguarda l'**efficacia**, non sono state trovate differenze significative comparando i vari effetti. Il task Checklist, che richiedeva soltanto di riconoscere se la percentuale rappresentata fosse maggiore o minore del 50% è risultato prevalentemente efficace. Diversi sono stati i risultati del task Slider, in cui era necessario inserire precisamente il valore rappresentato, difatti non è stato completato con successo dalla quasi metà degli utenti.

Si potrebbe ipotizzare che le visualizzazioni grafiche di percentuali proposte siano sufficientemente intuitive da rendere possibile individuarne il valore "a grandi linee", ma troppo poco chiare perché sia possibile riconoscerlo con precisione. È però importante ricordare che durante il test era possibile tornare indietro per osservare l'immagine di esempio raffigurante il 50% (e, esaminando anche brevemente i video dell'esecuzione dei task è chiaramente osservabile come ciò sia avvenuto nella stragrande maggioranza dei casi). È dunque possibile pensare che gli utenti abbiano completato con facilità il primo task solo grazie a questo aiuto.

Interessante a questo proposito è stato osservare i risultati del questionario compilato dagli stessi valutatori per il calcolo dell'OET: nonostante l'ormai buona conoscenza del sistema, si è registrato comunque un notevole tasso di errore. Questo potrebbe far supporre che effettivamente le visualizzazioni siano troppo vaghe per essere associate facilmente alla percentuale che rappresentano, senza una notevole dose di esercizio.

Potrebbe anche essere interessante analizzare la variazione dell'Error Rate al variare dell'ampiezza dell'intervallo all'interno del quale il task era considerato corretto.

Per quanto riguarda l'**efficienza**, inizialmente non parevano esserci considerazioni particolari da porre: i task venivano completati in alcuni minuti, un tempo adeguato considerata la scarsa familiarità col sistema. Un tempo che però, per i tre gruppi di effetti presi singolarmente, non presentava significative differenze con l'OET, che presupponeva invece una conoscenza più profonda del sistema. Si tratta di un risultato che contrasta con l'idea che fosse la bassa familiarità a rallentare gli utenti.

Possiamo però osservare che i p-value risultanti dai test effettuati non sono comunque particolarmente alti: sarebbe interessante svolgere un test ad una coda (supponendo che l'OET non sia maggiore dei tempi misurati sul campione) considerando un campione più ampio, per osservare eventuali differenze.

La misura della **soddisfazione**, invece, è risultata buona per quanto riguarda le misure di apprendibilità, attrattività, ed efficienza. Particolarmente inferiore è invece stato il risultato della controllabilità. Come rappresentato dai boxplots, però, per tutte le domande le risposte sono state prevalentemente positive.

Nella misurazione sussisteva il rischio di scontrarsi con il fatigue bias, il che avrebbe significato anche un elevato numero di risposte neutre (bias di tendenza centrale). Grazie agli accorgimenti presi però, questo non è accaduto: solo per 2 domande ("Noioso/Appassionante", "Conservativo/Non conservativo") più del 50% dei rispondenti ha dato risposta neutra. Per tutte le altre domande le risposte neutre sono state al di sotto del 50%.

Appendice

VALORI QUESTIONARIO TASK CHECKLIST							
ID	grup	BLUR		RUMORE		TRASPARENZA	
14	1	2	1	4	1	4	99
16	1	2	1	4	1	4	1
19	1	2	1	3	1	4	2
23	1	2	1	4	1	4	1
26	1	4	1	4	1	3	1
33	1	2	1	4	1	4	2
98	1	1	1	3	1	4	1
101	1	3	1	4	1	4	1
102	1	1	1	4	1	4	1
104	1	2	99	3	1	4	2
105	1	1	1	4	1	4	1
107	1	1	1	4	1	4	1
11	2	4	2	2	1	4	2
7	2	4	3	2	1	4	2
13	2	4	3	3	99	4	2
15	2	4	3	3	1	3	2
24	2	4	4	4	1	4	2
25	2	4	2	3	1	99	1
27	2	4	3	3	1	4	2
29	2	4	1	4	1	4	2
30	2	2	1	3	1	3	1
32	2	4	3	4	1	4	2
95	2	3	2	3	2	3	2
96	2	4	3	4	1	4	1
97	2	4	1	3	1	4	1
99	2	4	4	4	1	4	99
103	2	4	3	3	1	4	2
108	2	4	3	4	2	4	2
12	3	3	1	4	99	3	1
10	3	4	1	4	1	4	1
17	3	4	1	4	2	3	1
18	3	3	1	4	1	3	1
20	3	4	1	4	1	99	1
21	3	4	1	4	2	3	1
22	3	4	2	4	2	3	1
28	3	3	2	4	2	3	1
31	3	3	1	4	1	99	2
93	3	4	1	4	1	3	1
100	3	4	1	4	1	4	1
106	3	4	1	4	2	3	1

VALORI QUESTIONARIO TASK SLIDER

ID	G	BLUR						RUMORE						TRASPARENZA										
14	1			-40	36			-26	67			8	20			-10	71			19	50			-25
16	1	44		4	10			0	75			0	20			-10	80			10	30			-5
19	1	30	media	-10	15	media		-5	60	media		15	20	media		-10	70	media		20	40	media		-15
23	1	40	37,09	0	10	18,83		0	70	70,67		5	30	20,50		-20	100	76,58		-10	40	32,58		-15
26	1	74	dev standard	34	25	dev standard		-15	75	dev standard		0	25	dev standard		-15	76	dev standard		14	26	dev standard		-1
33	1	40	14,63	0	20	14,18		-10	85	7,09		-10	15	7,65		-5	75	8,32		15	38	7,90		-13
98	1	20		-20	10			0	60			15	10			0	75			15	25			0
101	1	30		-10	15			-5	70			5	20			-10	80			10	25			0
102	1	34		-6	16			-6	77			-2	14			-4	71			19	32			-7
104	1	41	valore corretto: 40	1	54	valore corretto: 10		-44	66	valore corretto: 75		9	38	valore corretto: 10		-28	78	valore corretto: 90		12	28	valore corretto: 25		-3
105	1	20		-20	0			10	73			2	14			-4	68			22	32			-7
107	1	35		-5	15			-5	70			5	20			-10	75			15	25			0
11	2	75		-15	47			13	50			10	30			-5	75			0	45			-5
7	2	76		-14	66			-6	43			17	30			-5	73			2	42			-2
13	2	75		-15	60			0	61			-1	36			-11	81			-6	43			-3
15	2	75		-15	54			6	58			2	25			0	61			14	41			-1
24	2	75	media	-15	63			-3	68			-8	23			2	75			0	39			1
25	2	75	71,25	-15	42	media		18	65	media		-5	26	media		-1	77	media		-2	33	media		7
27	2	79	dev standard	-11	59	49,31		1	55	59,44		5	45	33,88		-20	66	71,69		9	39	40,19		1
29	2	86	standard 14,90	-4	45	dev standard		15	60	dev standard		0	28	dev standard		-3	68	dev standard		7	40	dev standard		0
30	2	40		-50	25	13,86		35	59	10,08		1	30	12,54		-5	60	8,62		15	30	7,27		10
32	2	79		-11	54			6	69			-9	35			-10	70			5	46			-6
95	2	80		-10	45			15	60			0	36			-11	70			5	45			-5
96	2	30		-60	45			15	35			25	75			-50	90			-15	40			0
97	2	75	valore corretto: 90	-15	14	valore corretto: 60		46	65	valore corretto: 75		-5	33	valore corretto: 25		-8	61	valore corretto: 75		14	21	valore corretto: 40		19
99	2	80		-10	59			1	68			-8	25			0	65			10	50			-10
103	2	75		-15	60			0	60			0	25			0	70			5	40			0
108	2	65		-25	51			9	75			-15	40			-15	85			-10	49			-9
12	3	70		-5	34			-9	85			5				0	55			5	25			-15
10	3	60		-15	30			-5	75			15	40			0	55			5	20			-10
17	3	71	media	-4	21	media		4	75	media		15	35	media		5	65	media		-5	40	media		-30
18	3	68	67,00	-7	13	26,58		12	76	74,25		14	21	33,36		19	59	59,30		1	16	24,55		-6
20	3	74	dev standard	-1	15	dev standard		10	86	dev standard		4	35	dev standard		5	50	dev standard		10	20	dev standard		-10
21	3	70	4,13	-5	15	11,75		10	80	7,17		10	35	7,30		5	60	5,60		0	20	6,61		-10
22	3	70		-5	40			-15	74			16	40			0	60			0	25			-15
28	3	64		-11	48			-23	60			30	40			0	54			6	25			-15
31	3	62		-13	23			2	70			20	31			9				60	27			-17
93	3	65	valore corretto: 65	-10	15	valore corretto: 25		10	70	valore corretto: 90		20	30	valore corretto: 40		10	60	valore corretto: 60		0	30	valore corretto: 10		-20
100	3	65		-10	25			0	70			20	20			20	70			-10	20			-10
106	3	65		-10	40			-15	70			20	40			0	60			0	27			-17

TEMPI MISURATI SUL CAMPIONE (in secondi)			
BLUR	RUMORE	TRASPARENZA	Tempo totale*
32.59	51.85	402,09	497,2
44.03	35.03	467,13	561,75
100,71	39,92	56,97	215
76,35	306,04	101,57	501,84
34,95	369,93	33,34	451
186,46	81,13	93,22	394,29
72,65	193,44	53,55	319,64
439,28	46,93	78,99	565,2
70,47	44,9	176,76	292,13
44,34	34,88	91,16	170,38
40,1	32,86	200,88	273,84
54,17	165,72	49,11	269
259,22	88,89	39,27	403,15
149,25	89,82	69,65	338,32
58,25	320,32	43,7	443,12
74,88	57,27	42,78	193,77
104,77	34,73	47,63	206,94
75,96	41,7	45,26	178,99
31,7	48,22	94,62	191,07
414,13	50,89	67,65	555,45
442,05	29,23	60,87	543,62
42,51	57,93	112,96	230,53
67,36	41	326,28	434,64
69,58	132,77	76,49	278,84
37,53	45,83	1119,87	1203,23
71,39	196,44	173,18	441,01
72,41	662,59	149,96	884,96
35,47	514,69	53,78	603,94
68,92	70,95	120,59	275,87
60,84	233,17	69,89	384,89
57,91	77,17	48,56	197,53
45,59	30,83	223,77	312,58
52,42	123,29	73,68	266,53
88,78	112,23	124,06	339,27
87,23	113,22	66,73	282,23
43,05	160,57	31,73	275,56
25,68	58,65	44,2	141,08
56,68	60,22	1034,81	1151,71
52,15	69,35	1355,85	1477,35
44,89	1026,29	66,72	1137,9

*Il tempo totale è comprensivo anche del tempo impiegato dall'utente a rispondere alle domande di profilazione, se ha risposto, dunque in alcuni casi non sarà pari alla somma dei primi tre.

TEMPI MISURATI PER IL CALCOLO DELL'OET

BLUR	RUMORE	TRASPARENZA	Tempo totale
58,09	22,42	21,66	108,72
20,97	20,01	28,1	75,62
38,81	41	90,69	176,7
31,19	26,01	39,31	104,73
77,14	57,82	47,36	189,93
47,87	60,16	40,65	161,1
26,74	31,78	24,82	94,56
52,27	42,74	25,42	132,07
25,62	33,95	21,95	89,33

QUESTIONARIO DI VALUTAZIONE

Modulo di valutazione delle visualizzazioni alternative di percentuale.

Ciao! Dopo aver compilato il primo questionario, ti chiedo un rapido feedback su come è andata. Fra i cinque gradini presenti, che vanno dall'aggettivo più a destra fino al suo opposto sulla sinistra, scegli il più adatto a descrivere la tua esperienza.

*Campo obbligatorio

*
 1 2 3 4 5
 Fastidioso Piacevole

*
 1 2 3 4 5
 Incomprensibile Comprensibile

*
 1 2 3 4 5
 Creativo Privo di fantasia

*
 1 2 3 4 5
 Facile da apprendere Difficile da apprendere

1 2 3 4 5
 Noioso Appassionante

*
 1 2 3 4 5
 Non interessante Interessante

*
 1 2 3 4 5
 Imprevedibile Prevedibile

*
 1 2 3 4 5
 Veloce Lento

*
 1 2 3 4 5
 Originale Convenzionale

*
 1 2 3 4 5
 Ostruttivo Di supporto

*
 1 2 3 4 5
 Buono Scarso

*
 1 2 3 4 5
 Complicato Facile

*
 1 2 3 4 5
 Repellente Attraente

*
 1 2 3 4 5
 Usuale Moderno

*
 1 2 3 4 5
 Sgradevole Gradevole

*
 1 2 3 4 5
 Sicuro Insicuro

*
 1 2 3 4 5
 Attivante Soporifero

*
 1 2 3 4 5
 Conforme alle aspettative Non conforme alle aspettative

*
 1 2 3 4 5
 Inefficiente Efficiente

*
 1 2 3 4 5
 Chiaro Confuso

*
 1 2 3 4 5
 Non pragmatico Pragmatico

*
 1 2 3 4 5
 Ordinato Sovraccarico

*
 1 2 3 4 5
 Invitante Non invitante

*
 1 2 3 4 5
 Congeniale Non congeniale

*
 1 2 3 4 5
 Conservativo Non conservativo

Con quale probabilità consiglieresti l'utilizzo di questo tipo di immagini per visualizzare delle percentuali? *

1 2 3 4 5 6 7 8 9 10

AVANTI

Non inviare mai le password tramite Moduli Google

*
 1 2 3 4 5
 Buono Scarso

*
 1 2 3 4 5
 Complicato Facile

*
 1 2 3 4 5
 Repellente Attraente

*
 1 2 3 4 5
 Usuale Moderno

*
 1 2 3 4 5
 Sgradevole Gradevole

*
 1 2 3 4 5
 Sicuro Insicuro

*
 1 2 3 4 5
 Attivante Soporifero

*
 1 2 3 4 5
 Conforme alle aspettative Non conforme alle aspettative

*
 1 2 3 4 5
 Inefficiente Efficiente

*
 1 2 3 4 5
 Chiaro Confuso

*
 1 2 3 4 5
 Non pragmatico Pragmatico

*
 1 2 3 4 5
 Ordinato Sovraccarico

*
 1 2 3 4 5
 Invitante Non invitante

*
 1 2 3 4 5
 Congeniale Non congeniale

*
 1 2 3 4 5
 Conservativo Non conservativo

Con quale probabilità consiglieresti l'utilizzo di questo tipo di immagini per visualizzare delle percentuali? *

1 2 3 4 5 6 7 8 9 10

AVANTI

Non inviare mai le password tramite Moduli Google

RISPOSTE QUESTIONARIO DI VALUTAZIONE

Genere	Età	UEQ1	UEQ2	UEQ3	UEQ4	UEQ5	UEQ6	UEQ7	UEQ8	UEQ9	UEQ10	UEQ11	UEQ12	UEQ13	UEQ14	UEQ15	UEQ16	UEQ17	UEQ18	UEQ19	UEQ20	UEQ21	UEQ22	UEQ23	UEQ24	UEQ25	NPS
Uomo	20	5	5	2	1	4	5	3	1	1	5	1	4	4	4	5	1	1	3	4	1	2	1	1	2	4	7
Uomo	20	4	5	2	2	3	3	3	3	2	3	2	4	4	3	3	4	1	3	4	1	4	1	2	3	3	8
Donna	22	3	5	2	1	4	4	4	4	2	4	2	5	4	4	4	5	1	4	5	1	4	1	2	2	3	8
Donna	21	3	5	1	1	3	4	3	3	2	3	1	5	4	4	4	1	2	2	5	1	3	1	3	3	5	7
Donna	20	4	5	4	1	4	4	5	2	2	3	2	4	4	4	4	1	2	2	4	1	4	1	2	2	3	4
Uomo	19	4	3	4	3	3	4	2	2	2	4	2	3	4	4	4	2	2	4	2	1	4	1	2	2	3	7
Uomo	19	5	5	1	1	4	5	4	2	2	3	2	4	4	4	5	2	1	4	4	1	4	1	1	2	3	7
Donna	20	3	5	2	2	2	3	2	2	2	4	2	3	4	5	5	2	3	4	4	1	5	1	1	2	4	8
Uomo	22	4	5	3	2	2	3	4	1	3	4	2	5	3	4	4	1	3	3	2	2	3	3	3	3	3	4
Donna	22	3	5	1	1	3	3	3	1	1	2	2	2	4	5	4	1	1	4	4	2	3	1	2	2	4	8
Donna	20	3	5	1	1	3	5	2	1	1	4	2	5	4	4	4	2	3	2	4	1	4	1	2	1	2	8
Donna	26	5	5	1	1	4	5	3	1	1	5	1	3	4	4	5	1	1	3	4	1	5	1	1	1	3	8
Donna	25	5	5	1	1	4	4	4	1	1	4	1	5	4	4	5	2	3	3	5	1	4	1	1	1	3	7
Uomo	25	5	4	2	2	4	4	3	2	1	4	1	2	4	5	4	2	2	4	4	1	3	1	1	2	2	8
Uomo	24	4	4	3	2	3	3	5	1	2	3	3	5	3	4	4	2	3	2	4	1	3	2	3	3	3	3
Uomo	20	4	5	1	4	3	4	2	2	2	4	2	2	4	4	3	4	2	3	3	2	3	2	2	2	4	6
Uomo	24	5	3	2	2	3	4	3	1	2	3	3	4	4	4	5	2	2	3	3	3	3	1	2	2	3	7
Uomo	26	3	4	4	3	3	3	4	3	4	3	3	2	3	3	3	4	3	4	3	4	3	3	3	3	2	6
Uomo	21	5	5	2	1	3	3	1	3	1	3	2	5	3	5	5	4	2	3	5	1	3	1	2	3	2	6
Uomo	26	4	5	4	1	4	5	5	2	1	4	2	5	4	4	4	1	2	1	4	1	5	1	2	1	4	2
Uomo	22	5	5	1	1	5	5	1	1	1	5	1	3	5	5	5	1	1	1	5	1	5	1	1	1	3	10
Uomo	23	4	3	2	1	3	4	5	4	3	4	3	5	4	2	4	2	3	2	4	2	5	1	2	3	3	7
Donna	25	5	4	1	1	2	4	4	2	1	2	2	5	3	3	4	1	1	3	4	2	4	2	2	2	2	6
Uomo	25	3	5	3	1	3	4	4	1	2	4	2	5	3	3	3	2	2	2	4	1	4	1	3	2	3	7
Uomo	27	3	3	3	1	3	4	3	2	2	3	3	4	3	3	4	3	2	4	3	2	4	2	2	3	3	7
Donna	21	5	5	4	4	4	5	4	2	1	4	1	5	4	5	5	2	1	1	5	1	4	1	2	2	3	8
Donna	22	3	4	2	1	3	3	1	2	3	3	3	4	3	4	4	1	2	4	3	3	3	2	2	3	3	6
Donna	20	5	5	2	1	3	4	2	2	1	4	1	5	4	5	5	3	2	3	4	2	3	2	2	2	3	9
Donna	20	5	5	1	1	4	5	2	1	1	5	1	5	5	5	5	2	1	4	5	1	4	1	1	1	4	8
Donna	20	5	5	2	1	3	3	3	2	3	3	1	4	3	3	4	3	2	3	3	1	3	1	3	2	3	5
Uomo	20	4	3	2	2	4	4	5	1	2	3	2	4	4	4	4	3	1	2	3	2	2	1	2	2	3	7
Uomo	25	4	5	3	1	3	3	3	1	4	3	3	5	3	4	3	1	3	3	4	1	3	1	3	3	3	8
Donna	19	3	5	3	1	3	4	3	4	2	4	2	5	4	4	4	2	2	3	4	1	5	3	2	2	3	8
Uomo	20	4	5	2	2	4	3	3	2	2	3	2	4	3	3	4	2	2	3	3	2	4	2	2	3	3	4
Donna	19	3	5	2	1	3	4	2	1	2	5	2	5	4	4	3	3	3	4	4	2	3	2	2	2	3	7
Uomo	25	4	4	3	2	3	3	4	4	2	3	2	4	3	3	4	3	4	3	4	2	3	2	3	2	3	1
Uomo	24	3	5	3	1	3	4	5	3	2	3	2	3	4	3	3	2	2	3	3	2	3	2	2	3	3	7
Uomo	24	4	4	1	5	4	4	1	1	1	4	2	4	5	4	4	1	1	4	4	2	3	3	2	2	3	8
Uomo	26	2	5	1	3	4	5	2	3	1	3	1	3	4	5	3	2	1	3	4	2	4	2	1	1	3	9
Uomo	23	3	3	1	3	4	4	3	4	2	3	3	2	3	4	4	3	2	4	3	3	4	2	2	2	3	7

Biografia

¹ Wikipedia, "Infographics", acceduto il 10/07/2019, URL: <https://en.wikipedia.org/wiki/Infographic>

² Youtrend, "Europee 2019: un primo bilancio", acceduto il 10/07/2019,
URL: <https://www.youtrend.it/2019/05/27/elezioni-europee-2019-analisi-bilancio/>

³ Bootstrap, URL: <https://getbootstrap.com/>

⁴ CamanJS, URL: <http://camanjs.com/>