**UNIVERSITÀ DEGLI STUDI DI SALERNO**

  
  
  
Interazione Uomo Macchina e Usabilità del Software

**MedMinder**

**Assignment 3**

**Usability Testing**

***Francesca Festa -***

***Emanuele Gargiulo- 0522500718***

# Indice

[Indice 2](#_Toc10880077)

[Introduzione 3](#_Toc10880078)

[Definizione ISO dell’Usabilità 3](#_Toc10880079)

[Valutazione Empirica dell’usabilità 4](#_Toc10880080)

[Impostazione del testing per MedMinder 5](#_Toc10880081)

[Metodologia da Utilizzare 5](#_Toc10880082)

[Euristiche di Nielsen 6](#_Toc10880083)

[Requisiti di usabilità da testare 7](#_Toc10880084)

[Task da testare 7](#_Toc10880085)

[Scelta dei Partecipanti 7](#_Toc10880086)

[Questionario di background 8](#_Toc10880087)

[Questionario di Reazione 9](#_Toc10880088)

[Ipotesi da Testare 10](#_Toc10880089)

[Template per l’osservatore 10](#_Toc10880090)

[Validità Ecologica dello Studio 11](#_Toc10880091)

[Riferimenti //poi finiranno come ultima sezione 11](#_Toc10880092)

# Introduzione

## Definizione ISO dell’Usabilità

Secondo la definizione ufficiale dell’ISO 9241-11, l’usabilità è “la misura in cui un prodotto può essere utilizzato da utenti specifici per ottenere determinati obiettivi con efficacia, efficienza e soddisfazione.

* Efficacia: indica la capacità di raggiungere l'obiettivo prefissato. Si dice quindi che un prodotto software è efficace se permette di concludere un task che ci si era prefissati; nel caso in cui invece il task non viene completato con successo, l’efficacia si calcola considerando il numero di operazioni che si svolgono per portare a termine il task e la qualità del risultato ottenuto.
* Efficienza: valuta l’abilita di portare a termine un compito impiegando le risorse minime indispensabili. Principalmente per l’efficienza vengono considerati due fattori:  
     
  *tempo di utilizzo:* indica il tempo totale di utilizzo del sistema, per eseguire un task che ci si era prefissati. Un sistema risulta efficiente se il tempo di utilizzo risulta basso.  
    
  *spazio occupato:* indica la quantità di spazio in memoria occupata dal prodotto software. Minore è lo spazio occupato, maggiore risulta essere l’efficienza.  
    
  Per comprendere al meglio la differenza tra efficacia ed efficienza, occorre fare una distinzione. Considerati due sistemi A e B, se entrambi i sistemi portano a termine i task, sono considerati entrambi efficaci; se invece il sistema B porta a termine il compito in un minor tempo e con meno passi, è considerato efficiente rispetto al sistema A.
* Soddisfazione: processo volto a rilevare quanto è soddisfatto l’utente rispetto a un prodotto offerto. Il concetto di soddisfazione è correlato alle aspettative dell’utente e alla percezione della qualità del prodotto software offerto, al fine di migliorare il sistema avviando processi di rimodulazione dei servizi sulla base dei feedback raccolti. Il metodo più opportuno per misurare il grado di soddisfazione degli utenti è quello di raccogliere i loro pareri mediante questionari o interviste.

## Valutazione Empirica dell’usabilità

La valutazione empirica è lo standard utilizzato per effettuare la valutazione dell’usabilità, poiché permette di capire cosa succede quando il sistema è usato dagli utenti nel mondo reale.   
Questo metodo può risultare spesso rischioso perché se si attende troppo per far valutare il sistema all’utente, ossia se il sistema è già in una fase avanzata di progettazione, si rischia di dover rifare tutto daccapo.

Esistono diversi modi per portare avanti le valutazioni empiriche. Un primo metodo è quello dei *field studies* in cui si studia il sistema nell’ambito in cui deve essere effettivamente usato, raccogliendo quindi informazioni e classificandole secondo un grado di priorità in base ai problemi riscontrati. Questo procedimento è molto dispendioso in termini di tempo che economici. Per rendere più economico questo processo, si può pensare di far eseguire queste valutazioni agli utenti in laboratorio, assicurandosi che il risultato finale non sia condizionato dall’ambiente in cui si è svolto il test.

Un secondo metodo è quello delle *interviste retrospettive*, in cui si domanda all’utente di ricordare operazioni effettuate col sistema che risultano positive o troppo negative. Una pecca di questa tecnica è che si chiede all’utente di procedere mnemonicamente, il che porta spesso a problemi di inaffidabilità.

Lo scopo quindi del test di usabilità è quello di individuare eventuali criticità e colli di bottiglia dell’interfaccia o delle funzionalità del prodotto software, per poterli poi correggere. Questo tipo di test serve anche a capire come l’utente si muove e ragiona per capire quindi quali sono le eventuali difficoltà.  
Per comprendere il ragionamento dietro le azioni che l’utente esegue, gli si chiede di pensare ad alta voce, utilizzando quindi la tecnica del ***thinking aloud*** che prevede che l’utente spiega quali sono i propri obiettivi, le reazioni alle risposte del sistema e tutti i dubbi che gli vengono in mente durante l’interazione.

L’analisi dei dubbi espressi dall’utente in questa fase può aiutare a scoprire le cause e le conseguenze dei problemi di usabilità riscontrati dall’utente.

# Impostazione del testing per MedMinder

## Metodologia da Utilizzare

Avendo ricevuto il sistema da sestare in una fase di progettazione avanzata, o nella quale il sistema è addirittura già completo, la scelta di utilizzare test empirici per la valutazione è quasi obbligata.

La principale metodologia di valutazione adottata consisterà dunque nel far svolgere agli utenti una selezione di task principali in una situazione che simuli quanto più possibile il reale ambiente d’uso del sistema.

Mentre gli utenti svolgono tali task, i valutatori (in questo caso noi) prenderanno nota di tutte le azioni svolte dagli utenti – eventualmente, anche registrando video dell’interazione tra utente e sistema - al fine di poterle riesaminare per comprendere le motivazioni dietro i principali problemi incontrati nello svolgimento dei task.

Inoltre, agli utenti verrà chiesto di **pensare ad alta voce** (think aloud). In questo modo, oltre alle azioni effettivamente svolte dagli utenti, i valutatori avranno modo di esaminare anche i ragionamenti che hanno portato a tali azioni, ed eventuali dubbi che gli utenti si sono posti durante lo svolgimento dei task.

Oltre ai valutatori, con i quali l’utente non potrà interagire durante il test, è opportuno prevedere la presenza di un **facilitatore,** che possa fornire all’utente alcune informazioni o istruzioni di massima durante la prova, aiutando l’utente a superare situazioni in cui potrebbe non riuscire più ad andare avanti con i task.

I risultati ottenuti mediante questa valutazione empirica potranno essere eventualmente integrati con i risultati di una valutazione euristica, cioè verificando la corrispondenza del sistema ad alcune semplici ‘regole’ individuate da esperti del settore tra cui Nielsen, colui che ha elaborato le euristiche forse più celebri – di seguito riportate.

### Euristiche di Nielsen

I problemi d’usabilità nascono quando l’utente non riesce ad intuire le modalità di funzionamento del prodotto e deve necessariamente sforzarsi ad eseguire azioni molto lontane dalle sue aspettative e dalle sue normali modalità di esecuzione. Per comprendere al meglio quali siano gli aspetti positivi di un sistema e quali invece gli aspetti negativi, si possono seguire le 10 euristiche di Nielsen.

**1 – Visibilità dello stato del sistema**

* Il sistema deve sempre tenere informato l’utente su cosa sta facendo, fornendo un adeguato feedback in un tempo ragionevole.

**2 – Corrispondenza tra sistema e mondo reale**

* Il sistema deve parlare il linguaggio dell’utente, con parole, frasi e concetti a lui familiari.

**3 – Controllo e libertà**

* L’utente deve avere il controllo del contenuto informativo e deve potersi muoversi liberamente.

**4 – Consistenza e standard**

* L’utente deve aspettarsi che le convenzioni del sistema siano valide per tutta l’interfaccia.

**5 – Prevenzione dell’errore**

* Evitare di porre l’utente in situazione ambigue, critiche e che possono portare all’errore.
* Dare la possibilità di tornare indietro

**6 – Riconoscimento anziché ricordo**

* Produrre layout semplici e schematici
* Evitare che l’utente riscopra ogni volta l’interfaccia

**7 – Flessibilità d’uso**

* Offrire delle scorciatoie per i più esperti

**8 – Disegn e estetica minimalista**

* Dare maggior importanza al contenuto che all’estetica.
* Evitare di accentuare oggetti irrilevanti o raramente necessari (immagini grandi, etc.)
* Evitare che il contenuto informativo della pagina sia messo in secondo piano
* Evitare che l’utente si distragga o si confonda

**9 – Aiuto all’utente**

* I messaggi di errore devono essere espressi in linguaggio comprensibile e suggerire soluzioni
* Chiedere conferma per un’azione importante

**10- Documentazione**

* Anche se il sistema dovrebbe essere usabile senza documentazione è preferibile che essa sia disponibile

## Requisiti di usabilità da testare

In accordo con il progettista di MedMinder, i requisiti di usabilità da testare saranno:

* **Efficacia**: L’utente deve portare a termine tutti i task principali con scarse percentuali d’errore.
* **Efficienza**: L’utente deve portare a termine tutti i task principali velocemente.
* **Soddisfazione**: L’utente deve utilizzare il sistema con un sufficiente grado di soddisfazione.

Tali parametri verranno testati per i task principali dell’applicazione.

## Task da testare

* Aggiunta di un farmaco all’applicazione;
* Aggiunta di una visita all’applicazione;
* Visualizzazione del riepilogo delle medicine rimanenti;
* Visualizzazione del riepilogo delle visite mediche;
* Visualizzazione farmacie aperte in zona;
* Visualizzazione riepilogo generale.

# Scelta dei Partecipanti

La scelta dei partecipanti è cruciale affinché i risultati ottenuti dai test di valutazione siano attendibili.

Sarà utile suddividere gli utenti in due macrocategorie: quelli che hanno una certa familiarità con la tecnologia e quelli che invece non ne hanno.

Tutti gli utenti selezionati dovrebbero avere un potenziale interesse nell’utilizzo di un sistema come MedMinder.

I partecipanti dunque saranno selezionati tra colleghi dell’Università di Salerno - che rappresenteranno, verosimilmente, gli utenti con skill tecnologiche - e membri della famiglia dei valutatori, che rappresenteranno verosimilmente utenti con skill tecnologiche basilari o assenti.

Ad ogni modo, a ciascun utente verrà sottoposto un questionario di background per verificare tra le varie proprie il suo livello di conoscenza con la tecnologia – oltre che il livello di istruzione – al fine di poter analizzare i risultati ottenuti mettendoli potenzialmente in correlazione anche con questi aspetti.

## Questionario di background

Come accennato nella sezione precedente, a ciascuno dei partecipanti verrà consegnato un questionario di background da compilare, a fine di raccogliere maggiori informazioni che ci consentiranno poi di individuare eventuali correlazioni tra il background tecnologico/culturale ed i risultati dei test.

|  |  |  |  |
| --- | --- | --- | --- |
| Questionario di Background | | | |
| Nome | Cognome | Data | |
| Età: | Sesso: |
| Titolo di Studio | | | |
| Diploma Superiore | Laurea Triennale | Laurea Magistrale | Ambito: (Scientifico, Umanistico…) |
| Skill tecnologiche | | | |
| Quanto è familiare con l’utilizzo di uno smartphone? | | | |
| Pochissimo | Poco | Mediamente | Molto |

## Questionario di Reazione

In questa fase vengono proposte all’utente delle affermazioni relative all’uso del sistema e secondo la Scala di Likert l’utente dovrà esprimere quanto è d’accordo con tale affermazione in un range che va da “Fortemente in disaccordo” a “Fortemente in accordo”.   
Un questionario di questo tipo permette di valutare anche aspetti dell’usabilità che potrebbero non risultare da una valutazione puramente numerica come quella basata sulla percentuale di errori o sul tempo di esecuzione dei task.

In aggiunta a tale questionario, verranno poste all’utente delle **domande a risposta aperta** dandogli la possibilità di rispondere ad esse liberamente, per raccogliere ulteriori informazioni.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| L’applicazione è semplice da usare | | | | |
| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Mi è stato subito chiaro lo scopo dell’applicazione | | | | |
| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Penso che l’applicazione potrebbe effettivamente essermi utile | | | | |
| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Mi è stato subito chiaro lo scopo dell’applicazione | | | | |
| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Userò realmente l’applicazione, o la userei se ne avessi bisogno | | | | |
| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| I messaggi mostrati a schermo sono sempre chiari | | | | |
| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| È facile capire come svolgere ciascuna operazione | | | | |
| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sono soddisfatto della mia esperienza d’uso con l’applicazione | | | | |
| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |

Quali sono state le tre caratteristiche più interessanti del sistema? E le tre peggiori?

Cosa si potrebbe migliorare?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Ha altri consigli / feedback che desidererebbe fornire al team di sviluppo?

# Ipotesi da Testare

Verranno formulate delle ipotesi riguardo i risultati degli esperimenti, affermando che una variazione nelle variabili indipendenti *(IV)* causerà una variazione delle variabili dipendenti *(DV).*

Per fare ciò, sarà necessario invalidare l’ipotesi nulla – ossia, l’ipotesi per la quale le variabili dipendenti in realtà non vengono modificate in seguito a un cambiamento in quelle dipendenti.

Verrà adottato un testing **within groups:** Ciò significa che ciascun soggetto svolgerà i task sotto tutte le condizioni.   
Per limitare il più possibile i **learning bias**, si assegnerà a ciascun utente un ordine diverso di esecuzione dei task.

# Istruzioni sul testing e sui task

????

# Metodo di Raccolta

Ciascun valutatore avrà a disposizione il seguente template, che utilizzera per annotare vari aspetti relativi al test in uso

|  |  |  |  |
| --- | --- | --- | --- |
| UTENTE: | TASK: | Ora Inizio : | Ora Fine: |
|  | | | |
| Che schermata sta visualizzando l’utente? | |  | |
| Che dubbi sta esprimendo al riguardo? | |  | |
| Le ipotesi che sta formulando su come proseguire, sono corrette? | |  | |
| Quali difficoltà sta riscontrando? | |  | |
| Quali ambiguità del sistema hanno rischiato di confondere l’utente, se ve ne sono? | |  | |

In aggiunta a tale template, sarà utile registrare in maniera sincrona sia lo schermo del cellulare che il viso e le parole dell’utente. In tal modo, sovrapponendo le due registrazioni, sarà possibile esaminare le sue reazioni a ogni schermata presentata dal sistema, potendo anche ascoltare i ragionamenti dell’utente – dato che gli è stato chiesto di pensare ad alta voce.

# Validità Ecologica dello Studio

Dato che l’applicazione in uso verrà eseguita direttamente su uno smartphone, e che tutti gli smartphone degli ultimi anni dispongono di una fotocamera interna – oltre che della possibilità fornita tramite API di registrare lo schermo – per avere un’ambiente il più possibile simile all’ambiente di utilizzo reale, la registrazione delle schermate e dell’utente potrebbero avvenire semplicemente attivando sia la registrazione schermo del dispositivo che la registrazione del volto mediante fotocamera interna.

Tali schermate andrebbero in seguito sovrapposte per ottenere un video unico e sincrono dell’interazione tra utente e sistema.



Tuttavia, nel caso specifico del test portato avanti, non si disponeva della versione completamente implementata dell’applicazione né della possibilità di integrare in essa le funzionalità di registrazione mediante schermo e fotocamera.

Disponendo del prototipo in Marvel, è stato deciso di avviare quest’ultimo direttamente su smartphone – per ottenere un’esperienza d’uso il più realistica possibile e avere comunque la possibilità di registrare lo schermo – mentre l’utente verrà registrato mediante telecamere esterne.

I valutatori che dovranno osservare l’utente invece staranno nei suoi paraggi cercando di essere il più discreti possibili, ed intervenendo solo se strettamente necessario.

Se ciò dovesse avvenire, si prenderà prontamente nota di tale interferenza.



Figura 1 - Esempio di registrazione schermo (doppio click per avviare)

# Metriche

Le variabili dipendenti sono i parametri che vengono presi in considerazione quali misura dell’usabilità del sistema. Verranno utilizzate le seguenti metriche:

Per l’EFFICIENZA:

* Tempo di esecuzione del task

Per l’EFFICACIA

* Successo nello svolgimento del task

Per la SODDISFAZIONE

* Questionari somministrati agli utenti alla fine dello svolgimento di ciascun test, con domande aperte e affermazioni da valutare in un range da “Molto in disaccordo” a “molto d’accordo”

# Fase di Testing

## Fase Preliminare

Prima di iniziare il test vero è proprio, è stato opportuno spiegare a ciascuno dei partecipanti che tipo di applicazione avrebbero utilizzato, e quali sono i suoi scopi.

Bisogna inoltre sottolineare che è normale, in una fase di testing, individuare dei problemi: Non c’è niente di cui vergognarsi nell’ammettere di avere difficoltà, in quanto riscontrare tali difficoltà degli utenti aiuterà lo sviluppatore a migliorare l’usabilità dell’applicazione.

Per evitare agli utenti lo stress di affrontare il test senza aver mai visto il sistema, essi potranno farvi cinque minuti di pratica precedentemente al test vero e proprio, discutendo con il facilitatore eventuali dubbi generici sulle funzioni o sullo scopo del sistema.

Tuttavia, durante tali minuti di pratica, non sarà consentito fare domande specifiche e relative allo svolgimento dei task al facilitatore ai valutatori che presenzieranno: Se fossero questi ultimi a risolvere tutti i dubbi dell’utente, potenziali problemi di usabilità potrebbero non essere individuati e falsare i risultati del test.

## Svolgimento del Test

Durante lo svolgimento della prova, non sarà possibile alcuna interazione con i valutatori, che dovranno potersi concentrare pienamente sugli appunti da prendere.

L’interazione con il facilitatore dovrà a sua volta essere limitata al minimo necessario: egli fornirà rassicurazioni e incoraggerà l’utente al think-aloud, senza però interferire con lo svolgimento del task.

Conclusa l’interazione con il sistema, alla fine del test ad ogni utente sarà sottoposto il questionario di reazione costruito in precedenza, al fine di ottenere informazioni sul sistema che potrebbero non essere strettamente legate al successo o all’efficienza con la quale l’utente ha svolto i task.

# Analisi dei Risultati

Nella fase di analisi dei risultati, si raccoglie l’intero materiale ottenuto con il test – questionari degli utenti, appunti dei valutatori, video – e li si esamina al fine di ottenere dei dati che possano essere realmente utili nella valutazione dell’applicazione testata.

Tali risultati verranno poi forniti allo sviluppatore del sistema, che potrà verificare quali sono stati i problemi riscontrati con maggiore frequenza, o che hanno causato il maggior numero di problemi, ed intervenire per risolverlì, stabilendo quindi una scala di priorità degli aspetti da affrontare per migliorare l’applicazione.

Oltre che mediante l’analisi dei risultati ottenuti dagli utenti, la valutazione verterà anche sulla verifica da parte dei tester del rispetto di alcune euristiche.

## Valutazione Euristica

Come accennato nella sezione sulle metodologie da utilizzare per il testing, le euristiche da noi prese in esame per la valutazione sono le famose Euristiche di Nielsen.  
Per ciascuna di esse verrà evidenziato come viene implementata dal sistema, o se al contrario essa non trova corrispondenza nell’applicazione.

1. **Visibilità dello stato del sistema**
   1. Tale euristica sembra essere pienamente rispettata: Una volta selezionata ciascuna schermata, il nome della schermata sarà visualizzato sia in alto che sotto al bottone relativo al pannello selezionato.
2. **Corrispondenza tra sistema e mondo reale**
   1. L’utilizzo di icone che riprendono pienamente l’elemento che vogliono individuare (esempio: Casa per la schermata home, pillola per la schermata relativa ai farmaci..) aiuterà l’utente a comprendere immediatamente cosa troverà in ciascuna schermata.
   2. Riteniamo molto positiva ed intuitiva la scelta dello sviluppatore di visualizzare il livello di medicinale ancora disponibile in maniera grafica, mostrando cioè il flacone che si svuota man mano che il farmaco in questione viene consumato.
3. **Controllo e Libertà**
   1. Da tutte le schermate dell’applicazione risulta possibile raggiungere tutte le altre schermate principali, utilizzando un sistema di pannelli e bottoni comune a molte applicazioni moderne. Tale elemento è valutato quindi positivamente.
4. **Consistenza e standard**
   1. Tutte le schermate mostrano la stessa organizzazione grafica, con alcuni elementi che rimarranno visibili per qualunque schermata (l’intestazione e la barra con i bottoni di selezione pannello)
5. **Prevenzione dell’errore**
   1. Una volta intrapresa ciascuna azione, è disponibile una sola scelta (oltre a quella di tornare indietro). Sarà dunque molto difficile che l’utente possa sbagliare nello svolgimento di un task, una volta che l’avrà iniziato.
6. **Riconoscimento anziché ricordo**
   1. Le schermate sono tutte molto chiare, schematiche e rispondenti a quello che ci si aspetta oggigiorno da un’applicazione per smartphone. I periodi di apprendistato saranno molto brevi o nulli.
7. **Flessibilità d’uso**
   1. Il sistema non fornisce metodi d’uso alternativi o scorciatoie per utenti esperti. Sebbbene ciò contravvenga questa specifica euristica, è pur vero che lo scopo principale del sistema, cioè il ricordare l’assunzione di farmaci a persone anziane o che devono assumerne grande quantità, e l’utilizzo di una piattaforma mobile, lasciano poco spazio a modalità d’uso alternative, non essendo ad esempio possibile utilizzare comandi rapidi da tastiera.
8. **Design ed estetica minimalista**
   1. Il sistema è in linea con il *material design* definito da Google: Le schermate sono graficamente semplici e presentano pochi ‘fronzoli’ e nessun elemento non strettamente necessario all’uso.
9. **Aiuto all’utente**
   1. Nel prototipo fornitoci non risultano essere previsti prompt di conferma per le azioni principali, né sembrano essere presenti messaggi di errore. Sebbene data la semplicità d’uso dell’interfaccia ciò non rappresenti necessariamente un problema, sarebbe bene che lo sviluppatore riesaminasse questo aspetto.
10. **Documentazione**
    1. Allo stato attuale dello sviluppo non sembra essere presente una documentazione. Anche in questo caso, l’estrema semplicità d’uso potrebbe non renderla necessaria. Tuttavia, lo sviluppatore dovrebbe comunque prendere in considerazione l’ipotesi di rendere disponibile una documentazione per il sistema.

## Risultati dei test con gli utenti

# Riferimenti //poi finiranno come ultima sezione

1. http://www.far.unito.it/usabilita/Cap5.htm