Metodi di valutazione di usabilità

Esistono molti metodi di valutazione dell’usabilità, che si differenziano a seconda degli stili di valutazione adottati. La valutazione di un sistema si distingue in base a

* il luogo in cui avviene: laboratori di usabilità o nell’ambiente in cui gli utenti reali utilizzeranno il sistema
* chi la effettua: può essere effettuata da valutatori esperti e inoltre per alcuni metodi è necessario avere anche utenti reali.

Ci sono vari approcci alla valutazione (Interaction Design - Sharp et al. distinguono)

1. *Test di usabilità*: gli utenti provano un sistema o un prototipo, in genere nel laboratorio
2. *Studi sul campo*: conviene fare valutazioni nell'ambiente in cui gli utenti reali utilizzano il sistema
3. *Valutazione analitica*: gli utenti non sono coinvolti, in genere ci si basa su modelli formali applicabili a specifiche del sistema

Inoltre, è possibile raccogliere informazioni sull’usabilità tramite survey o procedere a valutazioni da parte di esperti in HCI, anche questa considerata come valutazione analitica.

I metodi usati da approcci diversi possono essere gli stessi ma variano le condizioni e i modi di applicazione.

I fattori considerare per scegliere il metodo di valutazione:

* *Fase del ciclo di sviluppo del sistema* in cui si sta facendo la valutazione (alcuni metodi si possono applicare anche per la valutazione di prototipi realizzati nella fase di progetto, altri nella fase di implementazione),
* *Stile di valutazione*: laboratorio di usabilità (ambiente controllato) piuttosto che nell’ambiente di utilizzo. Con prototipi iniziali si procede a una valutazione in laboratorio e man mano che si va avanti si valuta nell’ambiente di utilizzo per capire la vera portata del sistema.
* *Tipo di misure ricercate*: misure quantitative (es. numero di errori) oppure misure qualitative (l’intervista è una tecnica utilizzabile in fase di valutazione per acquisire informazioni sulla soddisfazione d’uso) che possono poi essere organizzate in strutture dati per derivare dati quantitativi. In ogni caso, la soddisfazione non può essere valutata con test ma tramite domande agli utenti.

Considerare anche le combinazione di fattori: i test considerano dimensioni (efficacia, efficienza) complementari rispetto alle domande agli utenti (soddisfazione).

* *Risorse richieste* per la valutazione: i metodi che richiedono solo il valutatore costano meno, mentre quelli che coinvolgono gli utenti richiedono di pagare anche, ad esempio, le società che li forniscono.

### Test di usabilità

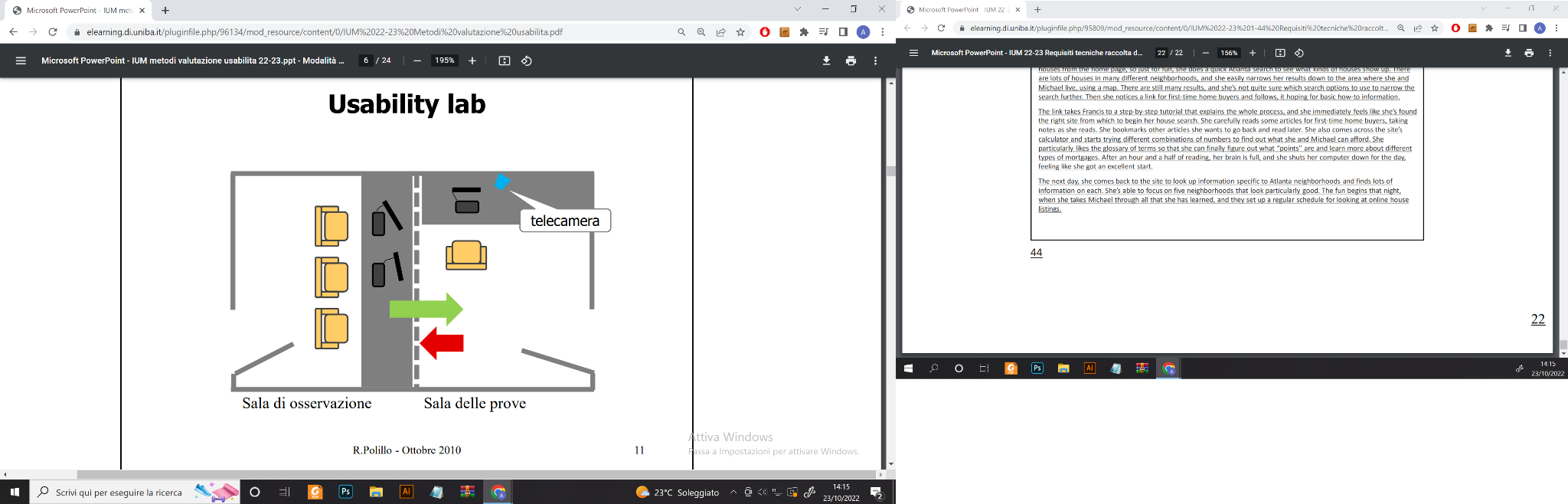
Utenti campione usano il sistema in un ambiente controllato: identificati pochi task significativi, si effettuano test sotto osservazione di esperti dell’usabilità che raccolgono dati, li analizzano e traggono conclusioni.

* Scopo: ricavare indicazioni concrete per il miglioramento del sistema
* Gli utenti scelti devono essere rappresentativi della categoria di utenti cui il sistema si rivolge
* Gli utenti devono tutti svolgere, separatamente, gli stessi compiti

Rispetto alle attività condotte dagli utenti durante le prove, i test di usabilità possono essere classificati in

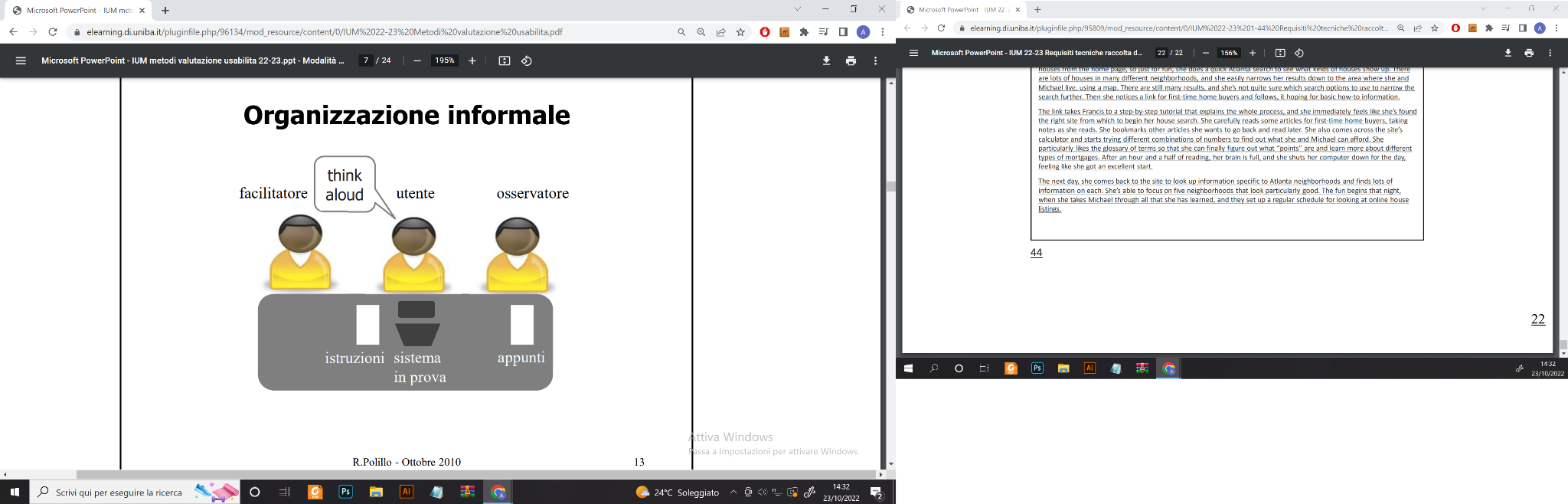
* **Test di compito**: viene fornita all’utente la lista di task specifici che permettono di esercitare le funzioni principali del sistema.
* Si può effettuare anche quando il sistema non è completamente sviluppato
* Obiettivo: capire quanto una funzionalità è effettivamente usabile
* **Test di scenario**: agli utenti viene indicato un obiettivo da raggiungere attraverso una serie di compiti elementari che non vengono indicati esplicitamente: l’utente dovrà quindi impostare una propria strategia di azioni (egli ha la descrizione di uno scenario, che rappresenta il contesto specifico in cui si deve muovere, e dovrà identificare task singoli da compiere per raggiungere l’obiettivo), compiendo un lavoro maggiore.
* Tipologia di test che si avvicina maggiormente all’uso specifico del sistema.
* Obiettivo: individuare eventuali carenze nell’impostazione della struttura complessiva dell’interazione o mancanze di funzionalità utili (anticiparli, per quanto possibile, durante lo svolgimento del progetto).

Le istruzioni agli utenti devono essere date per iscritto nel modo più chiaro possibile: solo in questo modo tutti i partecipanti al test si troveranno nelle medesime condizioni.

I test di usabilità si fanno in “laboratorio” (diversamente dagli studi sul campo) e possono essere organizzati in un **laboratorio di usabilità**, modalità che prevede costi alti, o in modo informale. 

Un laboratorio di usabilità cerca di riprodurre un ambiente simile a quello reale e prevede una sala delle prove, in cui è posto un finto specchio e tutti gli elementi per la ricostruzione della scena in cui l’utente utilizza abitualmente il sistema.

Tramite lo specchio, gli osservatori possono direttamente osservare l’utente dalla sala di osservazione; ci sono telecamere per riprendere l’espressione dell’utente nonché qualche sistema per registrare il prodotto sul dispositivo di utilizzo.



L’organizzazione informale, caratterizzata da costi notevolmente ridotti, prevede l’utilizzo di un tavolo qualunque su cui posizionare il dispositivo che utilizzerà l’utente e le istruzioni che egli dovrà seguire.

Saranno presenti due valutatori

* un facilitatore (o conduttore): deve facilitare lo svolgimento della prova, sollecitando il partecipante a verbalizzare tutte le sue azioni.
* un osservatore: osserva e prende nota del comportamento dell’utente. Egli deve conoscere bene il sistema, in modo tale che sia in grado di interpretare e valutare correttamente i comportamenti degli utenti.

Facilitatori e osservatori potranno essere esperti di usabilità o membri del team di progetto.

Il compito dell’osservatore è molto importante anche se il test viene registrato: l’esame di una registrazione può richiedere molto tempo e se l’osservatore ha già localizzato durante il test i problemi principali allora nell’analisi della registrazione ci si potrà concentrare sui punti critici, risparmiando molto tempo.

Viene chiesto all’utente di applicare il protocollo **thinking aloud** (di ricerca verbale, nato nell’ambito delle scienze sociali), ovvero egli dovrà pensare ad alta voce, specificando i ragionamenti che compie e le scelte sulle sue azioni: in questo modo si raccolgono i dati qualitativi sui problemi incontrati dal partecipante durante il test.

**Vantaggi**

* Il partecipante esprime a voce alta i pensieri, i dubbi, le percezioni man mano che esegue i task senza interrompere il flusso delle proprie azioni.
* Permette di avere un’idea più approfondita dello sforzo e delle problematiche dell’utente nell’uso del sistema
* Il protocollo è considerato più convincente nell’analisi di un task rispetto a osservazioni o interviste

**Svantaggi**

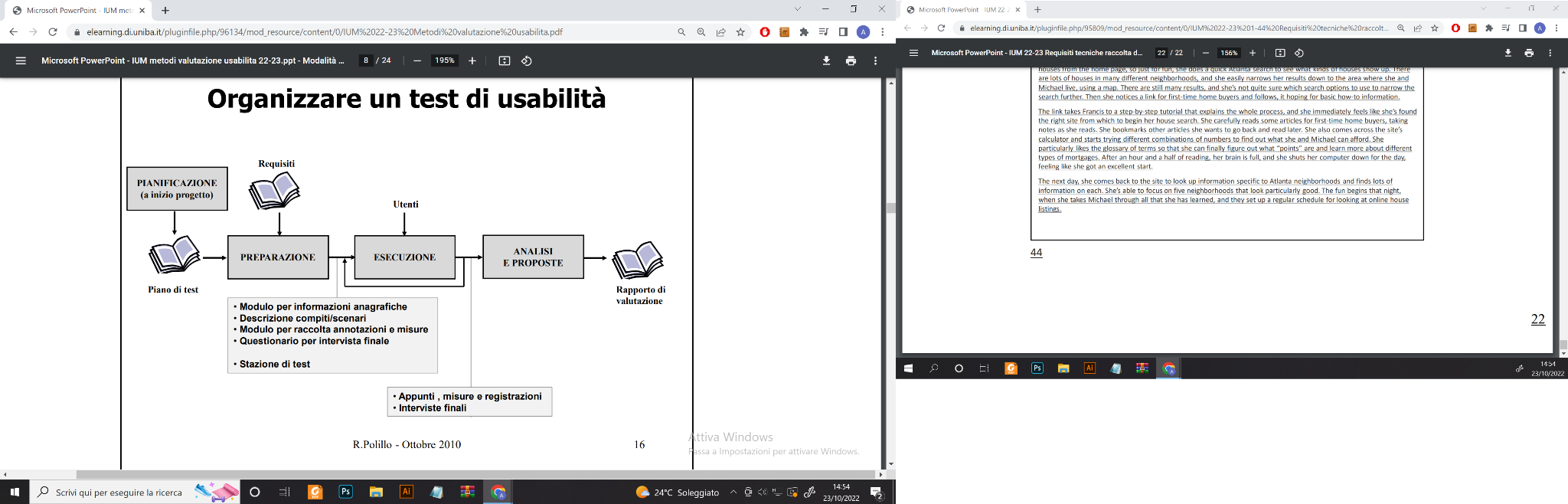
* il metodo di valutazione non è naturale, infatti il partecipante deve fare uno sforzo per verbalizzare tutto ciò che pensa (inizialmente l’utente pensa ad alta voce, poi inizia a perdere la concentrazione e rimane in silenzio). Per questo motivo il conduttore deve sollecitare l’utente a esplicitare ad alta voce
* cosa sta cercando di fare
* cosa vede sullo schermo
* come pensa di dover proseguire
* quali dubbi e difficoltà sta provando
* Il conduttore deve evitare di intervenire nel merito dei problemi verbalizzati (l’utente ha sempre ragione, infatti se fa qualcosa di sbagliato è perché il progettista non l'ha progettato in maniera adeguata), non deve porre domande dirette che possono guidare il partecipante né esprimere delusione o gioia per i comportamenti del partecipante.
* Alcuni partecipanti hanno difficoltà a pensare a voce alta, dunque con questi è bene non insistere nell’incoraggiamento.
* I dati prodotti non sono sempre quantificabili.

L’analisi del comportamento degli utenti non può essere condotta in tempo reale durante lo svolgimento del test ma dev’essere compiuta dopo, dato che le cause che portano alle difficoltà sperimentate possono non essere evidenti (serve riesaminare con attenzione la sequenza di azioni eseguite).

Oltre all’osservazione qualitativa dei comportamenti degli utenti è utile raccogliere anche delle *misure oggettive*: quelle più significative sono il tempo impiegato da ogni utente per l’esecuzione di ciascun compito, il tempo speso per correggere errori (misura dell’efficienza, in ogni caso si considerano relative statistiche) e il **tasso di successo** (*success rate*), ovvero la percentuale di compiti che ciascuno riesce a portare a termine (il calcolo può tener conto anche dei compiti eseguiti solo in parte). Si organizza una tabella utente-compito in cui vengono riportati gli esiti dei test.

**Organizzazione**

Un test di usabilità viene condotto in quattro fasi successive

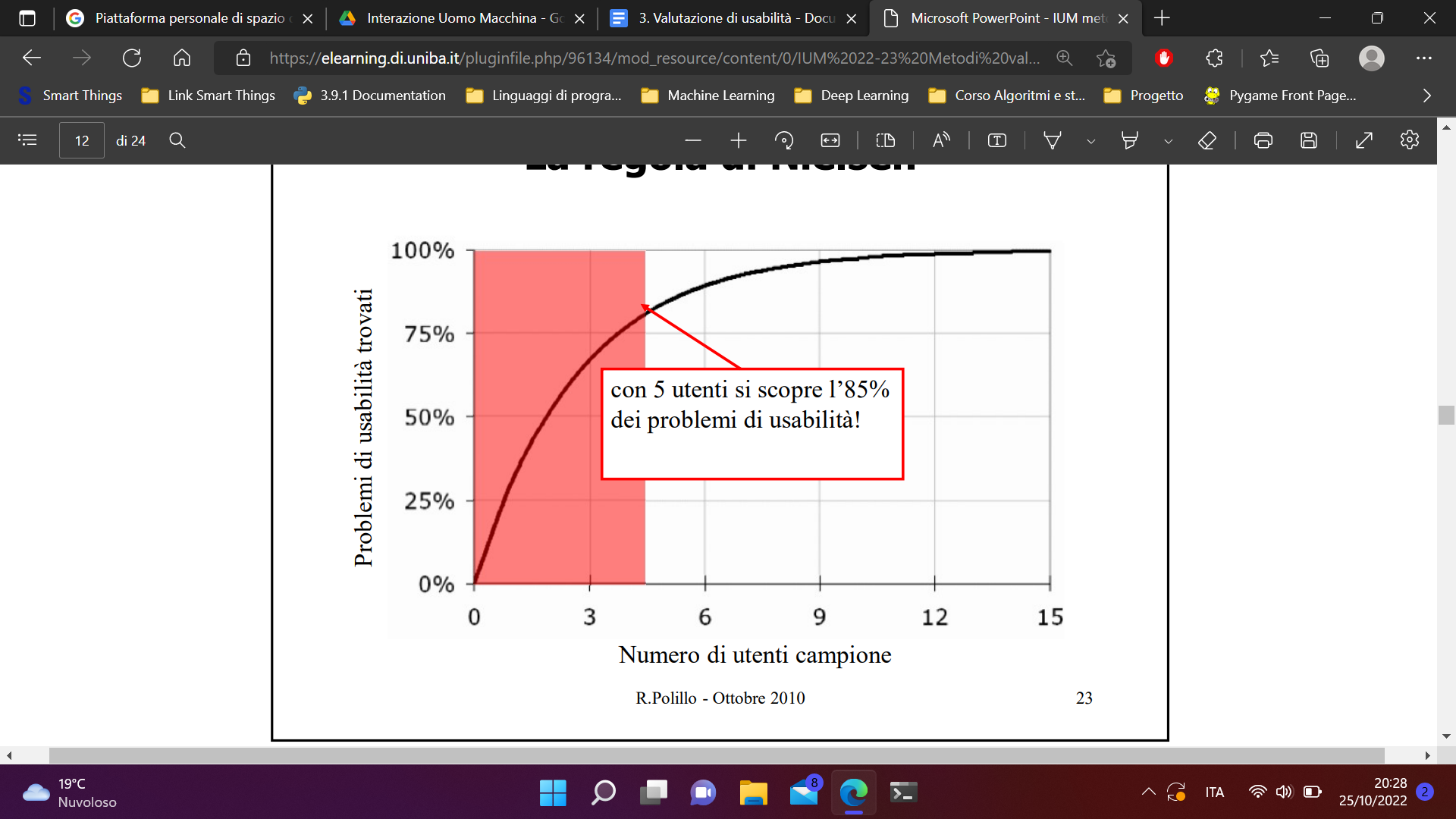
****

### Pianificazione

A inizio progetto, nella pianificazione delle HCI serve anche fare un piano dei test in cui si specifica quali parti del sistema devono essere valutate e come, come saranno realizzati i prototipi e come verrà eseguita la loro valutazione, quali saranno le interazioni con l’utente e come dovrà essere condotta l’analisi dei risultati.

Numero di utenti: dipende dagli obiettivi del test e dalla complessità del sistema (si rimanda a ciò che dice Nielsen)

* per check up rapido che avviene durante lo sviluppo in genere vengono coinvolti 5-7 utenti, assegnando 5-7 compiti ciascuno (20-40 min a utente)
* nella valutazione approfondita di un sistema complesso si arriva a 10-15 utenti impiegano 1-1:30 h per ciascuno.



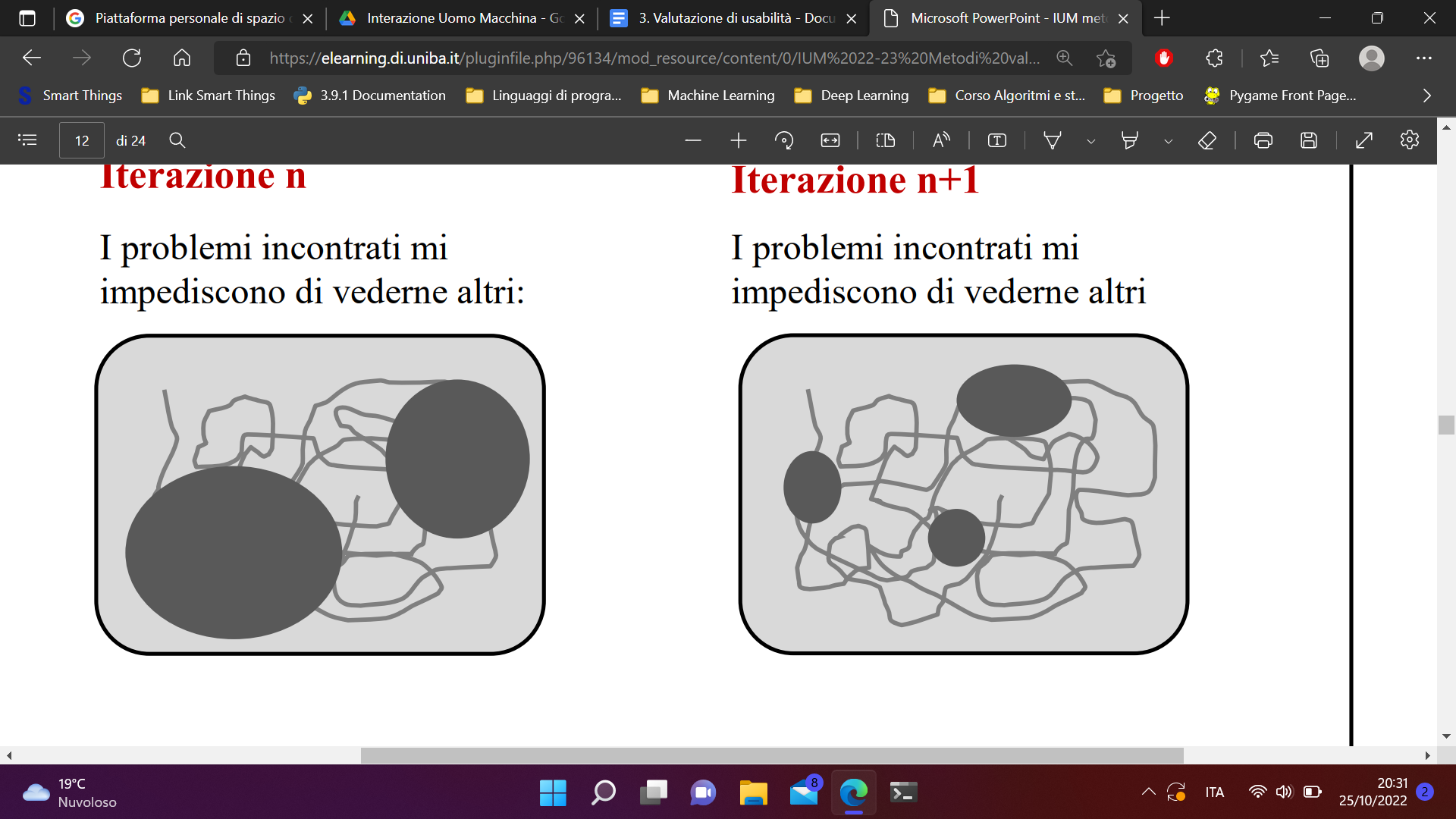
Nella valutazione in fase iniziale della progettazione, su prototipi dell’interfaccia utente, si verificano con l’utente le scelte migliori prima di procedere allo sviluppo

Per test che avvengono in questa fase, Jacob Nielsen afferma che dopo il quinto utente si sprecherà tempo osservando più volte gli stessi risultati: la curva relativa alla percentuale di problemi rilevati in relazione al numero di utenti intervistati assume un andamento logaritmico, quindi aumentare ulteriormente il numero di intervistati apporta benefici esigui.

* (3 → > 50%, 5 utenti → l’85% )

Per meglio specificare, a seconda dei loro obiettivi i test di usabilità possono essere classificati in

* **test formativi**, utilizzati durante il ciclo iterativo di progettazione per sottoporre i prototipi a prove d’uso con lo scopo di identificarne i difetti e migliorare l’usabilità → contribuiscono a “dare forma” al prodotto
  + Utili nelle fasi iniziali della progettazione
  + Mettono in luce i difetti macroscopici che richiedono una riprogettazione (parziale o totale) dell’interfaccia
  + Si verifica un **effetto di mascheramento**: i problemi riscontrati all’iterazione *n* impediscono di vederne altri essendo macroscopici. Una volta risolti, all’iterazione *n+1* se ne incontreranno altri.



* **test sommativi**: il loro obiettivo è una valutazione più complessiva del prodotto, in genere effettuata al termine del processo di progettazione e sviluppo.
  + Più completi di quelli formativi, valutano in modo sistematico pregi e difetti del prodotto o sue particolari caratteristiche.
  + Sono condotti di solito quando il prodotto è completamente funzionante e coinvolgono in genere un numero maggiore di utenti (10-15).

*Nota: test sommativi e formativi non classificano tipi particolari di test ma la denominazione dipende da quando il test è effettuato.*

Per poter interpretare l’esito di ciascun test, chi lo conduce dovrà conoscere il livello di esperienza d’uso di sistemi analoghi a quello in esame (per ciascun soggetto): nello specifico si deve caratterizzare l’utente che effettua il test sulla base della sua conoscenza del dominio applicativo del sistema e della familiarità con la tecnologia utilizzata (dimensioni indipendenti). Conviene scegliere utenti con caratteristiche diverse rispetto a queste dimensioni, per cui si suppone eserciteranno aspetti del sistema diversi.

* Evitare utenti che non hanno la minima conoscenza sul dominio → potrebbero incontrare difficoltà non dovute al sistema bensì alla loro poca dimestichezza, e i risultati delle prove sarebbero poco attendibili.

Decidere se effettuare test di compito o test di scenario costituisce una decisione critica

* effettuare un compromesso tra copertura delle situazioni possibili e tempo/risorse spese: in generale il test di compito risulta più semplice per valutatori e per gli utenti
* basarsi sulle priorità espresse nei requisiti

### Preparazione (materiali e ambiente di prova)

* scheda utente: include informazioni quali esperienza, conoscenza del sistema così come generalità dell’intervistato
* un modulo di consenso (liberatoria da far firmare all’utente)
* istruzioni per lo svolgimento delle prove, formulate in termini di una descrizione scritta dei compiti o scenari, da fornire all’utente per l’esecuzione (formulazione chiara e concisa)
* modulo per la raccolta di misure e osservazioni, da fornire all’osservatore: uno per ogni utente, l’osservatore dovrà riportare in modo separato le osservazioni oggettive e le sue considerazioni (es. tabella task-utente=
* questionario o domande per l’intervista finale degli utenti (aiuta a valutare la dimensione della soddisfazione dell’utente e possibili aspetti specifici dell’UX)

### Esecuzione

Non dura in genere più di qualche ora

* *Briefing agli utenti*: introduzione di benvenuto e di ringraziamento (dire di essere liberi, troveranno dei problemi e loro servono a identificarli).
  + Mettere gli utenti a proprio agio
  + chiarire che l’obiettivo della prova è valutare il sistema e non la capacità dell’utente
  + spiegare quali registrazione verranno fatte e la politica della privacy
* Fornito il materiale, effettuare il test conducendoli singolarmente: i valutatori dovranno interferire il meno possibile (solo il facilitatore dovrà rassicurare l’utente e incitarlo a proseguire con tranquillità secondo il protocollo thinking aloud).
* Al termine del test è utile intervistare gli utenti sull’esperienza fatta, utilizzando domande di un questionario appositamente predisposto, per sapere i punti di forza e di debolezza del sistema secondo l’opinione dell’utente o eventuali proposte per i progettisti che saranno scritte nel rapporto di valutazione.
  + es. questionario SUS, la cui validità è universalmente riconosciuta
  + In alternativa è possibile intervistare direttamente gli utenti o condurre un focus group finale.

*Nota: il facilitatore deve incoraggiare l’utente e non deve chiedere informazioni come “a cosa serve quel bottone”, “perché hai cliccato lì?”*

Pensare ad alta voce provoca interferenza condivisa, ovvero sovraccarico dell’utente che, concentrato sull’esecuzione del task, deve ricordarsi di pensare ad alta voce → serve a maggior ragione facilitatore

### Analisi dei risultati e proposte finali

Si effettua l’analisi dettagliata dei dati e della registrazione, elencando i singoli problemi e la loro gravità (A - alta, B - bassa, M - media).

Errori comuni:

1. limitarsi a riportare i giudizi espressi dagli utenti, i quali descrivono sensazioni di carattere generale ma non sono in grado di risalire con precisione a tutte le loro cause. Serve sempre compiere un’analisi dettagliata dei comportamenti.
2. Limitarsi a elencare poche difficoltà macroscopiche, senza andare in profondità. Serve elencare anche i problemi più piccoli individuati.

Il team di valutazione stende un elenco dei problemi riscontrati (descritti in modo circostanziato), assieme alle loro priorità nonché degli interventi suggeriti per migliorare il prodotto.

Si riassumono tutte le informazioni in un *rapporto di valutazione* (descrive i risultati dei test effettuati ma evidenzia anche il fatto che questi siano stati svolti in modo adeguato)

## Ispezione di usabilità

L’ispezione di usabilità è un’ulteriore tecnica di valutazione che utilizza un numero ridotto di risorse e sollecita l’osservazione critica da parte dell’esecutore.

I metodi d’ispezione dell’usabilità prendono l’idea dai metodi di ispezione del codice.

L’esperto effettua un’ispezione sistematica (di tipo visuale) delle schermate di un prodotto sulla base di euristiche, check-list e standard (gli utenti non sono coinvolti).

* Il valutatore annoterà i problemi rilevati su un modulo che riporterà ai progettisti del sistema. Essi valuteranno se e come apportare delle modifiche: la risoluzione di un certo problema potrebbe richiedere troppe risorse (ecco perché la valutazione va fatta nelle fasi iniziali del progetto).
* Il valutatore si cala nei panni di un utilizzatore che vede un sistema nuovo: segue compiti ponendo attenzione alla facilità di apprendimento, di esecuzione e cerca di capire cosa potrebbe dare problemi agli utenti.

L’ispezione va fatta in riferimento a principi di usabilità (es. euristiche di Nielsen) o riguardo la coerenza con una norma ISO/UNI o linee guida aziendali.

**Valutazione euristica**: tecnica di ispezione proposta da Jakob Nielsen (1993) che consiste nell’esame delle singole schermate sulla base di principi e linee guida (euristiche), eseguendo alcuni task rappresentativi.

L’utente esegue i compiti previsti per l’utilizzo del sistema, riflettendo su cosa non trova immediatamente chiaro e perché, capendo quale principio viene violato da quale elemento.

Deve proseguire in modo sistematico

*Parte statica*

* Si parte dalla schermata statica, controllando se gli elementi che appaiono sulla schermata sono significativi per l’utente nonché l’organizzazione del layout della stessa e che le informazioni siano chiare e non eccessive.
* In modo non frettoloso si passa alla schermata successiva dopo aver trovato i problemi nella precedente.

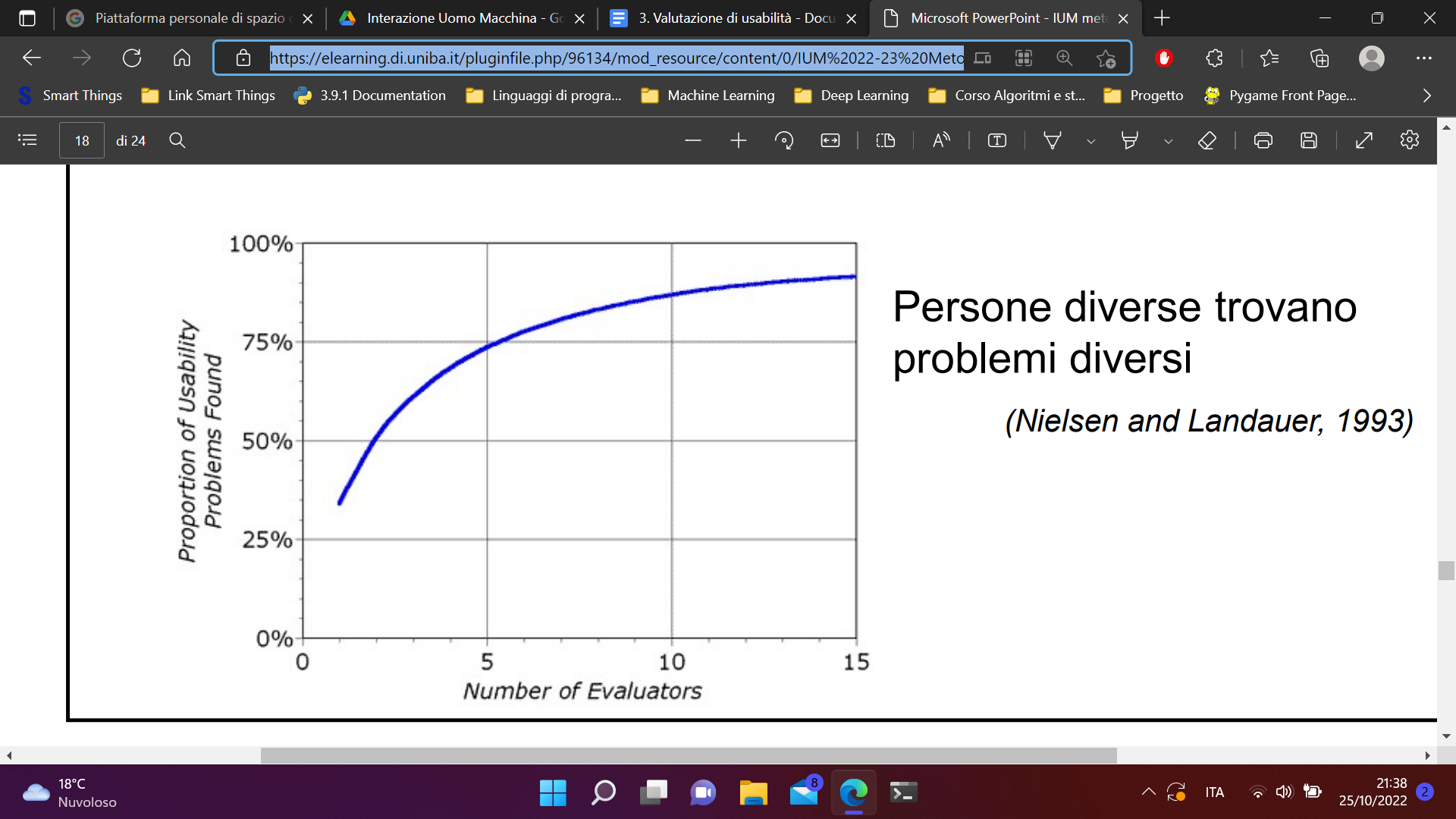
*Parte dinamica*

* Per eseguire un task l’interfaccia cambia, per cui serve ispezionare se i feedback sono appropriati e se il nuovo stato del sistema è reso evidente.
* stampare lista delle euristiche, eseguire un task (su una lista) e su un foglio annotare i problemi riscontrati.

Una persona sola può avere una valutazione soggettiva, per cui serve avere più esperti che la svolgono.

**Tabella di rilevazione dei problemi di usabilità**: si elencano una serie di problemi, specificando la loro locazione (in che pagina), descrizione e quale euristica violano (o più euristiche), accompagnando con una proposta di soluzione e indicando il grado di severità.

Questa tabella viene data ai progettisti, con i quali si discute anche delle possibili soluzioni.



Nielsen stesso chiama questo metodo “*discount usability*”, in quanto richiede poche risorse e presenta in generale un alto rapporto benefici-costi (consulenza di pochi valutatori, costo ancora più esiguo se li si ha in casa).

Anche in questo caso, l’andamento della percentuale di problemi di usabilità rilevati presenta un andamento logaritmico rispetto al numero di valutatori.

(2 valutatori 50%, 5 valutatori 75%).

**Problem rating**: dopo aver combinato le valutazioni di più valutatori in un’unica lista e rimosso i duplicati, ogni valutatore dà un voto a tutti i problemi riscontrati (es. scala da 0 a 4 per gravità crescente). Si effettua infine la media tra i voti, determinando le priorità per la correzione.

Questo processo è importante perché, mentre una valutazione individuale è un’opinione, il reale beneficio viene dato dalla combinazione di tutte le valutazioni.

*Giorgio Brajnik*: inserisce una serie di domande guida per aiutare valutatori poco esperti, che favoriscono un’analisi più approfondita dei problemi. Il valutatore deve però integrare la sua valutazione con altre domande, specifiche per il contesto dell’analisi (sito, utenza, scopo dell’indagine, …)

* Di risposta all’obiezione che le euristiche sono generali e forniscono pochi dettagli operativi (guidano poco il valutatore durante l’ispezione)

L’**esperienza del valutatore**:

* i novizi individuano 22% dei problemi
* esperti di usabilità 41%
* doppi esperti (usabilità+dominio) 60% dei problemi

Altre tecniche basate sull’utente

* tecniche di raccolta dati che pongono domande agli utenti possono essere usate anche per la valutazione (interviste individuali, questionari online, focus group); non possono però sostituire i test di usabilità
* esperimenti in laboratorio statisticamente significativi: richiedono una metodologia rigorosa, serve effettuare la scelta del campione e la definizione dell’ipotesi per poi condurre l’esperimento e utilizzare metodi statistici per l’analisi dei risultati.
* controllo di risposte fisiologiche: speciali apparecchiature permettono di effettuare il tracciamento delle risposte fisiologiche dell’utente durante l’interazione con il sistema
  + movimenti oculari
  + reazioni emotive (attività cardiaca, sudorazione, attività elettrica nel cervello, tono muscolare)

**Classificazione dei problemi**: qualunque sia la tecnica di valutazione utilizzata, i problemi individuati dovranno essere classificati sulla base della loro importanza

* Problema irrilevante → può non essere risolto
* Problema secondario → da risolvere con bassa priorità
* Problema rilevante → da risolvere con alta priorità
* Problema bloccante → deve essere necessariamente risolto prima del rilascio del sistema