

La valutazione del software. Le norme ISO/IEC 14598 e 15939

Gianluigi Raiss
Area Indirizzo, Supporto e Verifica PAC
Ufficio Pianificazione e Razionalizzazione
raiss@cnipa.it
<http://www.cnipa.gov.it>

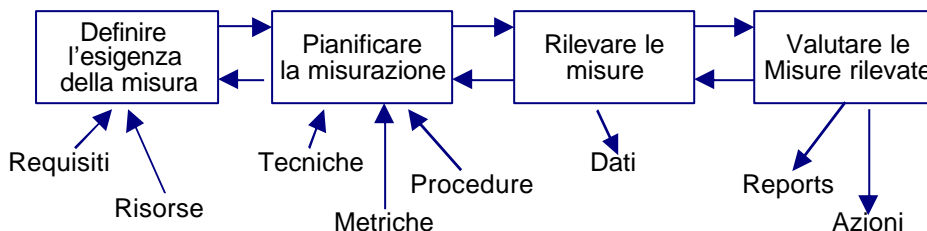
Concetti generali sulla misurazione del software. La norma ISO 15939.

Processo di misurazione del software

- Il processo di misurazione del software dispone di una norma ISO di riferimento:
 - ◆ ISO 15939 *Software Engineering – Software measurement process* (2002)
 - ? Il processo di misura è descritto nella norma come un ciclo iterativo che prevede la misurazione, la raccolta di feedback e l'utilizzo dei feedback per impostare azioni correttive per migliorare il processo di produzione (PDCA Cycle).
 - ? I termini utilizzati dalla norma si basano sullo *International vocabulary of basic and general terms in metrology* (ISO DGuide 99999:2004).

Il ciclo di misura

- Secondo la 15939, la misurazione del software è un processo iterativo PDCA, con 4 attività principali e 13 sottoattività.



La misurazione del software

- Nel ciclo di vita del software la misurazione serve a prevedere o stimare tempi di consegna, costo di lavorazione, qualità del prodotto.
 - ◆ Come in ogni altro settore ingegneristico, alle misure viene assegnato il compito di normalizzazione tra oggetti e fenomeni distinti, per un loro confronto o per effettuare correlazioni, al fine di valutare e prendere decisioni.
 - ◆ La tipicità (e “non fisicità” del software) rendono però in parte ambigue le misure in questo settore. Ciò non ha impedito all’ingegneria del software di proporre un vasto insieme di misure e di metodi per la misurazione.

Previsione e stima

- Nel ciclo di vita del software si può misurare lo stato raggiunto dal prodotto, nei suoi vari stadi di lavorazione, per 2 scopi principali:
 - ◆ la *previsione* delle caratteristiche che avrà il software in una fase del ciclo di vita diversa da quella in cui si effettua la valutazione (ad esempio, per impostare per tempo azioni correttive),
 - ◆ la *stima* delle caratteristiche possedute dal software, nella fase e nello stadio di sviluppo in cui si effettua la valutazione (ad esempio, per effettuare un collaudo od un test di accettazione).

Il concetto di misurazione

- Misurare vuol dire mappare un insieme di oggetti in un altro insieme di oggetti (numeri o simboli).
 - ◆ In altre parole, è una empirica ed oggettiva assegnazione di un numero (o simbolo) ad una entità al fine di caratterizzarne uno specifico *attributo*, usando una specifica *unità di misura*, una *scala di misura* e delle *regole* di conteggio (v. la definizione di *misurazione* in ISO 8402).
 - ◆ Per *entità* si intende qui uno stadio qualsiasi assunto dal software durante il processo di produzione (un *semilavorato*).

La misurazione - definizioni

- La ISO/IEC 14598-1 (che descrive il processo di valutazione del software), definisce la misurazione come: *“l’uso di una metrica per assegnare un valore (un numero od una categoria) su una scala predefinita”*.
- *“Il processo attraverso cui, secondo regole definite, simboli o numeri sono assegnati ad attributi di una entità”*. [N.Fenton, *Software metrics, a rigorous approach*, 1997].

Quando misurare il software....

- Vi sono diversi momenti che richiedono una valutazione sulle caratteristiche del software, durante tutto il ciclo di vita:
- in fase di progettazione:
 - ◆ per prevedere la facilità di eseguire interventi di manutenzione dopo il rilascio, per valutare l'impatto ed il costo che avranno nel contesto di utilizzo,
 - ◆ prevenire problemi nel software rilasciato in esercizio;
- in fase di collaudo e/o test:
 - ◆ per confrontare quanto fornito con le specifiche date,
 - ◆ scoprire problemi non considerati in fase di progettazione;

....quando misurare

- dopo il rilascio in esercizio:
 - ◆ per misurare l'impatto del prodotto sulla efficienza ed efficacia del lavoro svolto da chi ne fa uso,
 - ◆ confrontare le prestazioni del prodotto con quelle di altri prodotti comparabili,
 - ◆ individuare aree di possibile miglioramento,
 - ◆ decidere il momento del ritiro dalla produzione di una versione obsoleta.
- In sostanza, si misura per prendere decisioni ed agire.
 - ◆ *“Non puoi controllare ciò che non puoi misurare”* [T.De Marco, *Peopleware*]

Cosa misurare

- Cosa va misurato?
 - ◆ Un prodotto software è costituito non solo da codice sorgente, ma anche da documenti, regole, dati etc...
 - ◆ Inoltre, queste entità assumono nel ciclo di vita stadi diversi. Vanno effettuate quindi misure sia sui semilavorati che sugli stati finali delle entità che compongono un prodotto software.

Caratteristiche base delle misure

- Le misure utilizzate per valutare il software hanno alcune caratteristiche specifiche.
 - ◆ Scarso rispetto formale della teoria della misura
 - ◆ Focalizzate soprattutto nel migliorare la capacità di prevedere (e contenere) il costo del software
 - ◆ Stime soggette ad errori ed incertezze, in quanto applicate ad un sistema adattativo non lineare e complesso, in cui interagiscono molte variabili, in parte anche legate a fattori umani, organizzativi e gestionali

Misure interne (1/6)

- Misure interne del software, per fase del ciclo di vita
 - ◆ Misure nell'analisi dei requisiti – Utili per le implicazioni, a cascata, sul ciclo di vita del software e per stimare il costo del software
 - ✓ Numero dei requisiti
 - ✓ Function Points (i requisiti funzionali)
 - ✓ Volatilità dei requisiti

Misure interne (2/6)

- Misure interne del software, per fase del ciclo di vita
 - ◆ Misure nelle specifiche – Utili per le implicazioni, a cascata, sul ciclo di vita del software e per stimare il costo del software
 - ✓ Numero delle entità in un diagramma E/R
 - ✓ Numero blocchi in un digramma DFD
 - ✓ Numero degli Use Case Points (UML)

Misure interne (3/6)

- Misure interne del software, per fase del ciclo di vita
 - ◆ Misure nella progettazione – permettono di stimare il costo del software ma anche alcuni aspetti di qualità
 - ✓ Numero di moduli in uno schema di progettazione
 - ✓ Livello di coesione tra moduli
 - ✓ Livello di accoppiamento (coupling) tra moduli
 - ✓ Complessità di flusso (misura la quantità di informazioni in entrata ed uscita da una funzione [fan in e fan out] in ragione della lunghezza in Loc del modulo, v. Henry & Kafura, 1976)

Misure interne (4/6)

- Misure interne del software, per fase del ciclo di vita
 - ◆ Misure sul codice – permettono di stimare il costo del software ma anche alcuni aspetti di qualità. Sono le più diffuse e numerose.
 - ✓ Linee di codice (LOC o DSI) – usata dagli anni '60
 - ✓ Complessità ciclomatica (numero di Mc Cabe, 1976), è una misura del numero di nodi e archi del grafo del flusso di controllo
 - ✓ Misure di Halstead (1977) – si basano sulla combinazione di 5 elementi, derivati dalla teoria dell'informazione

Misure interne (5/6)

- Misure interne del software, per fase del ciclo di vita
 - ◆ Misure sul codice – cont'.
 - ✓ Misure di complessità (Van Emden, Harrison, loop connectedness, pubblicate dal 1977 in poi)
 - ✓ Misure di Troy (1981) – insieme di 24 misure per valutare la modularità, lunghezza, complessità, coesione, accoppiamento del software
 - ✓ Misure object oriented – misurano ereditarietà, polimorfismo, incapsulamento etc.. (da fine anni 80)
 - ✓ Misure di coesione funzionale – misurano le istanze di definizione e utilizzo di variabili e costanti (1994)

Misure interne (6/6)

- Misure interne del software, per fase del ciclo di vita
 - ◆ Misure sul test – Permettono di verificare la qualità del test effettuato sul software
 - ✓ Livello di copertura di istruzioni, dei rami, dei percorsi base

Misure esterne

- Misure esterne del prodotto software
 - ◆ Sono tra le più diffuse ma anche tra le meno scientificamente comprese, in quanto hanno diverse sfaccettature e potenziali interpretazioni
 - ✓ Metriche di Gilb, 1988
 - ✓ Modello di qualità di Bohem
 - ✓ Modello di qualità di Mc Call
 - ✓ ISO/IEC 9126-1 (in pubblicazione ISO 25000)

Le metriche - definizioni

- Per effettuare le misure occorre definire una unità di misura, una scala di misura ed il metodo utilizzato per la misura. Questi elementi sono definiti nel loro complesso dalla ISO/IEC 9126 come “metrica”.
- ◆ Una metrica è quindi: *“un metodo di misura associato ad una scala di valutazione”*.
- ◆ Secondo il glossario IEEE una metrica è *“la misura quantitativa del grado con cui una entità possiede un determinato attributo”*.
- ◆ *Un modello che descrive una variabile dipendente in funzione di una variabile indipendente...*

Tipologie di metriche

- Le metriche possono essere classificate secondo diversi criteri:
 - ◆ metriche *elementari*, in grado di rappresentare immediatamente un dato fenomeno (ad es., il metro lineare per le lunghezze, i secondi per gli intervalli di tempo etc...),
 - ◆ metriche *derivate*, che possono essere definite come funzioni (matematiche) ricavate da operazioni su altre misure (ad es. le funzioni statistiche, la media, la varianza, il minimo, il massimo etc..).

Metriche elementari e derivate

- ◆ Esempi di metriche elementari sono il tempo espresso in secondi (s) od anni solari (a), l'impegno espresso come *full time equivalent* (FTE) o giorni-persona (gp), il costo espresso in euro, il volume del software espresso in LOC o function points (FP).
- ◆ Esempi di metriche derivate sono la produttività, espressa come rapporto tra il volume di codice sviluppato e l'impegno (FP/gp), la difettosità, espressa come rapporto tra il numero di difetti ed il volume di codice sviluppato (n° / LOC).

Metriche dirette ed indirette....

- Rispetto alla capacità di misurare con immediatezza le caratteristiche del software, si possono distinguere metriche *dirette* ed *indirette*.
 - ◆ Le metriche dirette sono quelle rilevabili sul software senza dover considerare la influenza di fattori esterni, come l'ambiente (hardware e software) nel quale il software "gira" od il comportamento e le caratteristiche degli utenti etc... Esempio di queste misure sono quelle rilevabili leggendo il codice sorgente e quelle rilevabili da una lettura dei documenti di analisi e/o specifica.

...metriche dirette ed indirette

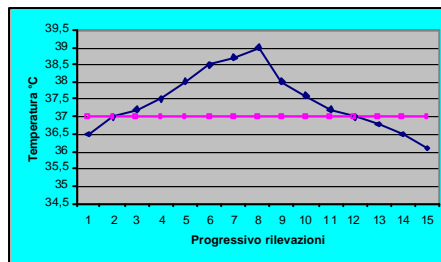
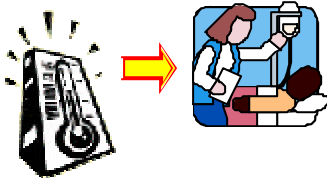
- ◆ Le metriche indirette combinano nella misurazione delle caratteristiche del prodotto anche misure relative ad elementi del contesto.
 - ? Per esempio, le misure relative ai tempi di risposta di un prodotto software non dipendono solo dal comportamento del prodotto in sé, ma anche da quello dell'ambiente operativo nel quale gira il software (hardware e software di base, rete...). I tempi di risposta mostrati da un prodotto software contengono anche i tempi di risposta di questi componenti.
 - ? Le misure di usabilità di un prodotto contengono anche quanto si riferisce alla capacità degli utenti di operare sul prodotto, e quindi la loro formazione, le attitudini, competenze, motivazioni etc...

Gli indicatori

- Alcune misure possono essere stimate a partire da altre misure, dette indicatori.
 - ◆ Per es., il tempo di risposta del software non è misurabile quando il software è ancora allo stato di *non eseguibile*. Ma può essere utilizzata la misura della semplice lunghezza del codice stesso (numero di stmts o di Loc o di FP) come indicatore di quello che sarà il tempo di risposta del prodotto nell'ambiente d'uso.
 - ◆ Gli indicatori “predicono” il comportamento del prodotto, in determinate condizioni, o stimano l'andamento di un fenomeno a partire dalla misura di un altro, con il quale è in correlazione.

Esempi di indicatori

- La ISO/IEC 14598 definisce indicatore: “*una misura che può essere utilizzata per stimare o predire un'altra misura*”.
- ◆ Temperatura di un corpo: misura ricavabile misurando la lunghezza di una barra di mercurio in un termometro.



Uso degli indicatori

■ Previsione

- ◆ Caratteristica da misurare: capacità di apprendimento,
- ◆ Indicatore: presenza ai corsi,
- ◆ Metrica: contare le presenze dello studente alle lezioni, rispetto alle lezioni svolte,
- ◆ Valore di soglia: >90% = buona
tra 70 e 89% = discreta
<70% = insufficiente

■ Stima

- ◆ Indicatore: correttezza delle risposte all'esame,
- ◆ Metrica: n° risposte esatte / n° domande,
- ◆ Valore di soglia: > 90% buona....

Proprietà degli indicatori

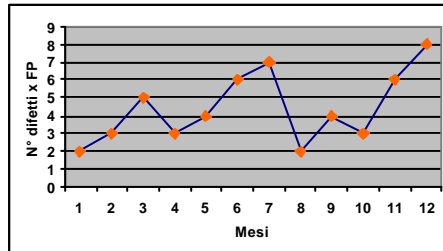
■ Gli indicatori devono rispettare la teoria della rappresentazione:

- ◆ le misure rilevate devono conservare le relazioni esistenti nel mondo reale anche nella loro conversione in valori di un sistema simbolico, in modo che osservando ciò che accade ai numeri si può predire ciò che accade nella realtà,
- ◆ al variare di un fenomeno, deve potersi osservare un analogo variare della misura utilizzata per osservare il fenomeno.

Misure di andamenti (trend)

- Un tipo particolare di metriche mette in relazione due grandezze al variare di una di esse.
 - ◆ Ad es. il variare della difettosità del software in funzione della distanza temporale dal suo rilascio in esercizio,
 - ◆ il variare dei costi di sviluppo del software nel tempo.

Questo tipo di metriche serve in genere a misurare comportamenti nel tempo (trend).



Misure di dimensione, tempo, occorrenza

- Le misure applicabili al software possono essere classificate in base alla grandezza che misurano. Si distinguono così misure di *dimensione*, *tempo*, *occorrenze*.
 - ◆ Misure di dimensione – Il software, come ogni altra entità, può essere rappresentato attraverso varie dimensioni (così come ad esempio un solido può essere rappresentato dalla massa, dal volume, dalla superficie etc...).
 - ◆ Le più utilizzate nel caso del software sono la dimensione funzionale (unità di misura i FP) e la lunghezza (unità di misura le LOC).

Misure di dimensione - utilizzo

- LOC e FP vengono utilizzati per misurare indirettamente altre grandezze rilevanti in un progetto:
 - ◆ la probabilità di presenza di un errore nel programma (cresce con la dimensione),
 - ◆ il tempo necessario a produrlo,
 - ◆ la produttività che può avere un programmatore,
 - ◆ il costo di produrre un programma,
 - ◆ La difficoltà di apportare modifiche al programma (in manutenzione) dopo il rilascio in esercizio.

Misure di tempo....

- L'unità di misura in questo caso può essere il “real time” (secondi, minuti etc..) ma anche il “clock time” del processore (secondi di CPU) od il “running time” di un componente del sistema. Tra le misure vi sono:
 - ◆ *tempo operativo*, in genere usato come base per valutare la affidabilità di un prodotto. Si possono misurare lo “elapsed time” (distanza di un evento da un preciso riferimento temporale), od il tempo operativo di un componente (il tempo in cui effettivamente “funziona”),
 - ◆ *tempo di esecuzione*, il tempo necessario ad un componente ad effettuare una data operazione,

....misure di tempo

- ◆ *tempo utente*, il periodo speso da un utente per completare un compito svolto utilizzando un supporto informatico. Si possono misurare la durata di una sessione (ad esempio per capire il comportamento di un utente di fronte ad un sistema interattivo cui richiede dei servizi attraverso dei comandi da immettere tramite una interfaccia), o la durata di una singola operazione ,
- ◆ *tempo produttivo*, depurato dalle interruzioni, richiesto ad un individuo (o ad un team di individui) per svolgere un determinato compito (ad esempio, per sviluppare un certo numero di linee di codice),
- ◆ *l'intervallo di tempo tra due eventi consecutivi* in un periodo di osservazione (ad esempio, l'intervallo tra due fermi consecutivi di un componente).

Misure di occorrenza....

- Le misure di occorrenza contano il numero di volte che un evento si verifica in un periodo di osservazione. Tra queste misure vi sono:
 - ◆ il numero di difetti rilevati in un test, ispezione etc...od anche il numero di tipologie di difetti rilevate (ad esempio, presenza/assenza di funzionalità richieste, incorretta operatività di funzionalità, inadeguatezza delle funzionalità a servire le esigenze dell'utente etc...),
 - ◆ le misure di complessità strutturale, come quelle di McCabe (il numero cicломatico),
 - ◆ il numero di inconsistenze rispetto ad un riferimento (ad esempio le non conformità rispetto ad una norma, ad un modello di requisiti, ad un formato di dati etc...),

....misure di occorrenza

- ◆ il numero di modifiche apportate ad un programma (od il numero di LOC/FP/STMTS modificati),
- ◆ il numero di errori nel disegno architeturale e tecnico, i requisiti utente non correttamente compresi correttamente nella fase di analisi etc..
- ◆ il numero di azioni umane necessarie ad interagire con un prodotto informatico per eseguire un compito, misure che riguardano in particolare la usabilità (ad esempio, il numero di tasti da premere per effettuare un'operazione, il numero di movimenti dell'occhio necessari etc..),
- ◆ un punteggio assegnato da un utente tramite un questionario.

Combinare misure per valutare meglio....

Combinando le tipologie di misure di Dimensione (Size), Tempo (Time), Occorrenza (Count) si possono rappresentare i principali fenomeni connessi al software.

<i>Combinazione</i>	<i>Adatta a rappresentare</i>	<i>Esempio</i>
Tempo / Dimensione	L'efficienza o lo sforzo	- Tempo di esecuzione di uno step di un programma - Tempo di risoluzione di una anomalia per stmts di source code
Occorrenza / Dimensione	La densità di un fenomeno	- Densità di difetti od errori - Copertura dei test
Dimensione / Dimensione	La incidenza di un fenomeno	- Numero ciclomatico per modulo - Percentile di moduli con alta densità di errori
Dimensione / Tempo	Andamento nel tempo	- Volume di sw sviluppato per mese

...combinare le misure

<i>Combinazione</i>	<i>Adatta a rappresentare</i>	<i>Esempio</i>
Occorrenza / Tempo	La frequenza	- Numero di errori umani commessi nell'interagire con il sw in un mese - N° di transazioni per secondo - N° di LOC modificate per mese
Tempo / Tempo	La incidenza	- % di disponibilità effettiva di un componente rispetto al tempo operativo
Dimensione / Occorrenza	La estensione	- N° medio di cammini linearmente indipendenti testati in un test
Tempo / Occorrenza	Intervalli di tempo	- Tempo tra due errori consecutivi (di un componente hw o sw, od umani) - Tempo necessario a riparare un difetto rilevato
Occorrenza / Occorrenza	La incidenza	- N° di operazioni umane compiute correttamente nell'utilizzo del software

Definizione di scala di misura

- Si definisce *scala di misura* l'insieme di:
 - ◆ un sistema di relazioni empiriche S
 - ◆ un sistema di relazioni numeriche R
 - ◆ un mapping tra S ed R
- Poiché il numero di possibili insiemi di relazioni empiriche e numeriche è pressoché infinito, infinite sono le scale di misura utilizzabili.
 - ◆ Si definisce trasformazione di scala una funzione che trasforma una scala in un'altra. Il *tipo* di scala è determinato dall'insieme di trasformazioni possibili per quella scala.

Tipi di scale di misura....

- *Nominale* - le misure sono semplicemente collocate in uno schema di classificazione. Non vi è una relazione di ordinamento tra entità appartenenti a classi diverse (sono *classi di equivalenza*).
 - ◆ Un esempio di classificazione nominale può essere quella relativa ai possibili tipi di errori di programmazione (di sintassi, semantici etc. senza dare una scala di gravità degli errori).
 - ◆ Se si utilizza questo tipo di scala, calcolare valori medi ha significato solo se si vuole verificare la frequenza con cui certe misure si collocano in date classi. Altre operazioni tra le misure non hanno significato.

....scale di misura....

- *Ordinale* - le misure sono collocate in una relazione che prevede una possibilità di loro ordinamento, ma non la possibilità di comparare la differenza tra le diverse classi di misure (quanto distano tra loro).
 - ◆ Ad esempio, le malfunzioni possono essere classificate per *media-grave-disastrosa*, per cui possiamo dire che *disastrosa* è peggio di *grave*, ma senza poter dire di quanto in termini analitici.
 - ◆ Se si utilizza questo tipo di scala, calcolare la mediana ha significato solo se si vuole verificare la frequenza con cui certe misure si collocano in date classi. Altre operazioni possibili sono la determinazione di "maggiore di" o "minore di", il calcolo dei percentili.

...scale di misura....

- *Intervallare* - In questa scala la differenza tra due misure ha un significato empirico e permette di ordinare le misure.
 - ◆ Un esempio di misura che utilizza questa scala è la temperatura, od il voto ad un esame su una scala 18-30, in sostituzione di *sufficiente-buono-ottimo*.
 - ◆ Possono essere fatte anche operazioni di addizione e sottrazione tra le misure. Si possono calcolare la media aritmetica, la deviazione standard.

...scale di misura

- *A rapporti* - In questa scala la differenza tra due misure ed anche la proporzione tra due misure permette di ordinarle.
 - Si possono effettuare anche moltiplicazioni e divisioni tra misure. Si possono calcolare la media armonica e geometrica, la variazione percentuale, l'indice di correlazione.
 - Esempi di questo tipo di scala sono gli intervalli di tempo, le lunghezze, i pesi.
- *Assoluta* - In questa scala si contano le occorrenze effettive delle entità.
 - Ad esempio, le LOC che costituiscono un programma.

Sinottico scale di misura (1)

TIPO DI SCALA	TRASFORMAZIONI AMMISSIBILI	OPERAZIONI EMPIRICHE DI BASE	ALCUNI INDICI STATISTICI APPROPRIATI	TEST STATISTICI APPROPRIATI	ESEMPI
NOMINALE	qualsiasi trasformazione uno-a-uno	determinazione di uguaglianza	Moda Frequenza	non parametrico	etichettare classificare entità
ORDINALE	$M(x) \geq M(y)$ implica che $M'(x) \geq M'(y)$ [funzione crescente strettamente monotona]	come la precedente, più la determinazione di "maggiore di" e "minore di"	Mediana Percentili Spearman r Kendall W Kendall T	non parametrico	preferenza ordinamento di entità
INTERVALLARE	$M' = aM + n$ ($a > 0$) [trasformata lineare positiva]	come la precedente, più addizioni e sottrazioni	Media Aritmetica Deviazione standard Correlazione Pearson Correlazione multipla	non parametrico	temperature in gradi Fahrenheit o Celsius date di calendario orario
RATIO (DI RAPPORTI)	$M' = aM$ ($a > 0$) [trasformazione di similarità]	come la precedente, più moltiplicazioni e divisioni	Media Geometrica Media Armonica Coefficiente di variazione Variazione percentuale Indice di correlazione	non parametrico e parametrico	intervalli di tempo temperature Kelvin lunghezze
ASSOLUTA	$M' = M$ [identità]				conteggio di entità

Scelta della scala di misura

- La scelta della scala dipende dall'attributo che si vuole misurare. La scala scelta deve corrispondere ad un effettivo sistema di relazioni valido per l'attributo. I valori numerici associati alle misurazioni dell'attributo devono corrispondere ad effettive relazioni empiriche fra le entità misurate.
- ◆ Ad esempio, se per un determinato prodotto non è possibile stabilire che è "affidabile" in maniera esattamente doppia o tripla etc.. di un altro, non ha senso utilizzare una scala diversa da quella ordinale.



La ISO/IEC 14598

Contenuti della ISO/IEC 14598

- La norma ISO/IEC 14598, *Information Technology, Software Product Evaluation* descrive il processo di valutazione della qualità del software, in accordo con la norma ISO/IEC 9126.
- ◆ La norma è composta di 6 parti, pubblicate tra il 1998 ed il 2001 (ed in parte recepite anche in Italia da UNI). E' in corso di revisione come ISO 25040.
- ◆ La norma è utilizzabile per valutare il software in fase di sua acquisizione (il modello di valutazione è nella parte 4), durante lo sviluppo (parte 3), in fase di assessment di terza parte (la parte 5). Nella parte 2 sono indicazioni per pianificare e gestire una valutazione, nella parte 6 un esempio di documentazione da produrre a valle della valutazione. La parte 1 è introduttiva.

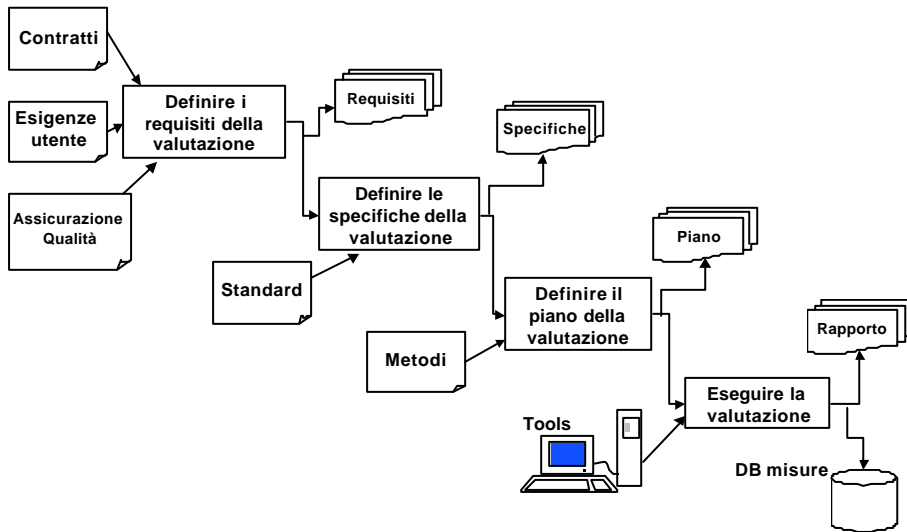
Applicabilità della ISO/IEC 14598

- Il modello di valutazione proposto nella norma si intende applicabile sia ai prodotti *commercial-off-the-shelf*, sia a quelli sviluppati su commessa in base a specifici requisiti utente (*custom software* o software *ad hoc*), sia alle personalizzazioni di prodotti COTS.
- Attività e *task* di valutazione definiti nella norma sono coerenti con la ISO/IEC 12207 (che definisce processi, attività e task del ciclo di vita del software, incluso quindi quanto va fatto per assicurarne e valutarne la qualità).

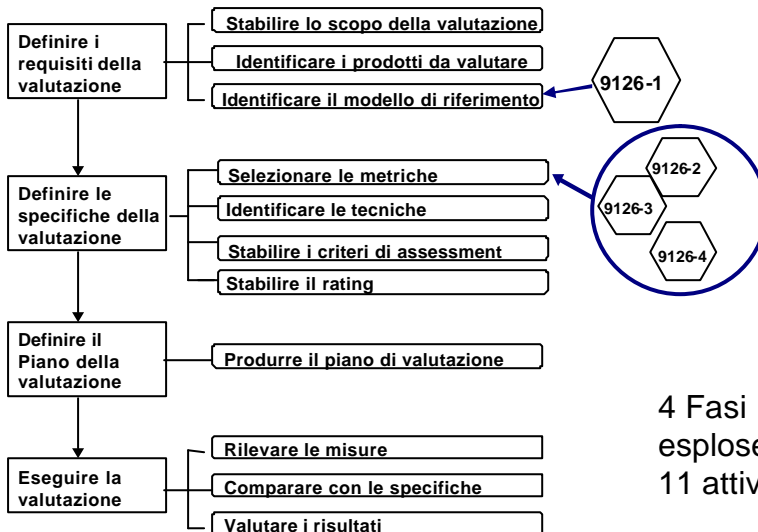
Alcune definizioni

- Dalla ISO /IEC 14598-4:
 - ◆ Valutazione della qualità: “*esame sistematico del livello con il quale una entità risponde ai requisiti stabiliti*”
 - ◆ Verifica: “*Conferma attraverso un esame e la rilevazione di evidenze oggettive che uno specifico requisito è stato rispettato*”
 - ◆ Validazione: “*Conferma attraverso un esame e la rilevazione di evidenze oggettive che uno specifico requisito relativo ad un particolare obiettivo d'utilizzo è stato rispettato*”.

Schema del processo di valutazione



Fasi/Attività del processo di valutazione



4 Fasi esplose in 11 attività

Requisiti della valutazione

- Nei requisiti della valutazione vanno definiti questi elementi:
 - ◆ la descrizione dello scopo della valutazione,
 - ◆ la identificazione dei prodotti da valutare, incluse le interfacce, i dati etc..),
 - ◆ le attese dell'utente in termini di livello di caratteristiche di qualità che deve possedere il prodotto espresse utilizzando il modello di riferimento ISO 9126-1 (profilo di qualità atteso).

Definire lo scopo della valutazione....

- Le valutazioni possono essere svolte sul prodotto finito, oppure nelle fasi intermedie del ciclo di sviluppo, sui semilavorati.
- Lo scopo della valutazione del prodotto finito può essere:
 - ◆ decidere sulla accettabilità del prodotto (anche decidere se risponde ai requisiti di una specifica normativa),
 - ◆ confrontare il prodotto con altri concorrenti, in fase di selezione o di valutazione di una eventuale sostituzione,
 - ◆ decidere quando modificare un prodotto, durante l'esercizio, per adeguarlo a nuove esigenze, migliorarne le caratteristiche, correggere problemi.

...definire lo scopo della valutazione

- Lo scopo della valutazione nelle fasi intermedie dello sviluppo può essere:
 - ◆ decidere sull'accettazione di stati intermedi del prodotto,
 - ◆ decidere se la fase è da considerarsi completata e se si può passare alla successiva,
 - ◆ prevedere la qualità del prodotto finito in base a quella degli stati intermedi del prodotto,
 - ◆ rilevare informazioni sugli stati intermedi del prodotto al fine di migliorare la gestione del processo.

Ruoli vs scopo della valutazione....

- ◆ l'**Acquirente** deve definire i requisiti di qualità "esterni" ed "in uso" che desidera nel prodotto (cfr. ISO/IEC 9126-1, 2 e 4) e valutare le soluzioni proposte dagli sviluppatori (*suppliers*) nelle offerte tecniche;
- ◆ il **Fornitore** deve valutare i propri prodotti per assicurarsi che corrispondano alle aspettative dell'utente (ad es. espresse in un contratto) e non abbiano caratteristiche inferiori a quanto comparabile (per costo e funzionalità) sul mercato;
- ◆ lo **Sviluppatore** di prodotti custom deve tradurre le aspettative qualitative del Committente in specifiche tecniche e requisiti di qualità interni (cfr. ISO/IEC 9126-3);

...ruoli vs scopo della valutazione

- ◆ il **Collaudatore** deve accertare la conformità delle caratteristiche del prodotto consegnato rispetto ai requisiti del Committente;
- ◆ il **Manutentore** del prodotto deve verificare che non si siano degradate le caratteristiche qualitative del prodotto sottoposto a manutenzione, dopo il rilascio in esercizio;
- ◆ il **Gestore** del prodotto (post rilascio) deve assicurarsi che le caratteristiche desiderate siano raggiunte e mantenute nelle condizioni operative del prodotto, e fornire informazioni utili alla pianificazione di eventuali interventi migliorativi.

Identificare l'oggetto da valutare

- Un prodotto software assume diverse forme (stadi o semilavorati) nelle diverse fasi del ciclo di vita.
 - ◆ All'inizio sarà valutabile la sola documentazione di analisi, al termine del ciclo di sviluppo sarà possibile indagare su documentazione, codice sorgente, manuali d'uso, manutenzione ed operativi, dati associati.
- Lo stadio di lavorazione determina il tipo di verifica:
 - ◆ la qualità in uso può essere valutata solo quando il prodotto è operativo ed utilizzato dall'utente.
 - ◆ Le caratteristiche interne ed esterne possono essere valutate anche in stadi intermedi: in questo caso, le misure serviranno a predire la qualità del prodotto finito ed in uso e ad attivare eventuali azioni correttive/preventive.

Stadio lavorazione vs verifiche

- In funzione dello stadio di lavorazione si possono valutare diverse caratteristiche (o predirle):
 - ◆ nei documenti progettuali, la correttezza e validità formale, ovvero che non siano riscontrabili errori concettuali, sia rispettata la congruenza e consistenza interna ed esterna dei documenti, la correttezza semantica, la chiarezza nelle descrizioni, il possesso del livello adeguato di strutturazione delle descrizioni, rispetto al loro possibile utilizzo nelle fasi successive del ciclo di vita del progetto, la comprensibilità, l'accuratezza;
 - ◆ nel prodotto in fase di collaudo, tutte le caratteristiche 9126, (interne, esterne ed in uso);
 - ◆ nel prodotto in esercizio, la usabilità, la qualità in uso.

Definire il modello di riferimento

- Il modello base di riferimento per definire le caratteristiche da valutare è quello ISO/IEC 9126-1.
 - ◆ In genere, le caratteristiche definite in questo modello non sono tutte insieme richieste in un prodotto. La scelta di quali considerare rilevanti nello specifico prodotto, ed a quale livello di possesso, va fatta secondo le esigenze dell'utente e del contesto. Allo scopo, va rilevato il *profilo di qualità atteso* dall'utente.
 - ◆ Si costruisce questo profilo intervistando un campione rappresentativo di utenti, con tecniche appropriate (ad es. *brainstorming*, *checklist*, simulazioni con prototipi a diversi stadi di completezza).

Rilevare la qualità attesa...

- Per definire il profilo di qualità atteso va costruito un questionario da sottoporre agli utenti.

		Funzionalità		Affidabilità		Usabilità		Efficienza		Mantenibilità		Portabilità			
Domande	Caratt e- ristica	Possibile risposta	Valut az.	F(P)	peso F	A(P)	peso A	U(P)	peso U	E(P)	peso E	M(P)	peso M	P(P)	peso P
.....	F, U, A	alto = 3 medio = 2 basso = 1 nullo = 0			1		2		2		0		0		0
.....	F, E	alto = 3 medio = 2 basso = 1 nullo = 0			1		0		0		2		0		0
.....	F, M, P	alto = 3 medio = 2 basso = 1 nullo = 0			1		0		0		0		1		3

....rilevare la qualità attesa

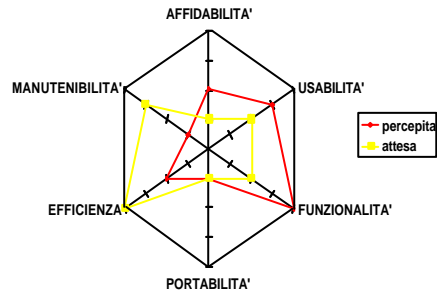
- Di norma le domande sono pesate.

		Funzionalità		Affidabilità		Usabilità		Efficienza		Mantenibilità		Portabilità			
Domande	Caratt e- ristica	Possibile risposta	Valut. medi a	F(P)	peso F	A(P)	peso A	U(P)	peso U	E(P)	peso E	M(P)	peso M	P(P)	peso P
.....	F, U, A	alto = 3 medio = 2 basso = 1 nullo = 0	1,8	1,8	1	3,6	2	3,6	2	0	0	0	0	0	0
.....	F, E	alto = 3 medio = 2 basso = 1 nullo = 0	1,4	1,4	1	0	0	0	0	2,8	2	0	0	0	0
.....	F, M, P	alto = 3 medio = 2 basso = 1 nullo = 0	2,7	2,7	1	0	0	0	0	0	0	2,7	1	8,1	3
Somma val.pesati					5,9		3,6		3,6		2,8		2,7		8,1
Val. Normalizzati					2,0		1,8		1,8		1,4		2,7		2,7

Completare il questionario

- Al profilo risultante possono essere applicati dei fattori di correzione (come viene ad esempio fatto per stimare i Function Points dopo aver calcolato quelli *unadjusted*, secondo il metodo IFPUG 4),

I fattori di aggiustamento tengono conto di ulteriori elementi di contesto, non noti all'utente intervistata (ad esempio il già citato rischio, il budget ed il tempo disponibile, od altri vincoli di contesto e di sicurezza).



Considerare la criticità

- Un fattore di peso da considerare è il livello di criticità del prodotto nel suo contesto di utilizzo. Allo scopo, è utile definire una tabella di correlazione tra livelli qualitativi e classi di rischio conseguenti a possibili malfunzioni nel prodotto, facendo riferimento ad aspetti economici, di sicurezza, di danni all'ambiente od alle persone.

Livello 0	livello 1	livello 2	livello 3	livello 4
Nessun rischio	danni non irreversibili a cose	danni irreversibili a cose	possibili danni a persone	a rischio la vita di persone

Specificare il metodo di valutazione...

- Nelle specifiche di valutazione vanno definiti:
 - ◆ le fonti da cui rilevare le misure necessarie alla valutazione;
 - ◆ le tecniche che verranno usate per effettuare le valutazioni;
 - ◆ le metriche che verranno utilizzate (v. ISO/IEC 9126-2, 3 e 4) per rilevare misure sul software;
 - ◆ le elaborazioni che verranno effettuate sui dati rilevati (aggregazioni, operazioni, arrotondamenti etc...);

....specificare il metodo di valutazione...

- ◆ la consistenza del campione da esaminare; le valutazioni possono essere svolte con differenti livelli di copertura delle caratteristiche del prodotto e con diverse accuratezze.
 - ? Il valutatore si deve porre il problema di scegliere il campione confacente al budget disponibile, al contesto ed alla criticità del prodotto.
 - ? Questa scelta potrebbe essere in contrasto con la possibilità di indagare su tutti gli aspetti desiderati.

....specificare il metodo di valutazione

- ◆ l'ambiente di svolgimento delle verifiche. Tale ambiente (strumenti, risorse, condizioni di svolgimento), andrebbe definito specificando le condizioni che identificano almeno tre differenti ambienti, "basso utilizzo", "a regime", in situazione "di stress";
- ◆ la strategia delle verifiche, che devono permettere valutazioni su funzioni "elementari", in cui ogni funzione è valutata isolatamente dal resto delle componenti del prodotto, senza tener conto delle relazioni che esistono tra loro, e test sulle sequenze di funzioni (attivate in sequenza).

Fonti della valutazione

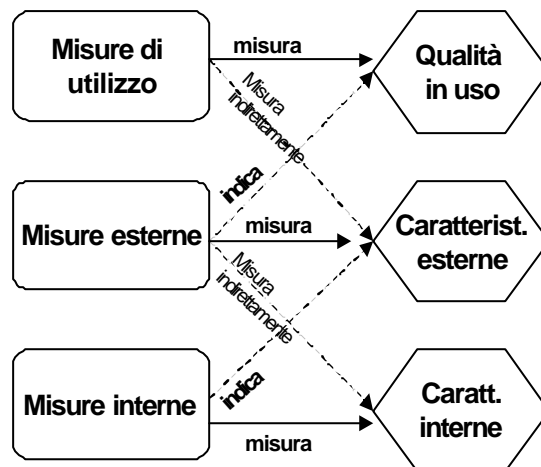
- Le fonti della valutazione sono sostanzialmente due:
 - ◆ l'entità stessa in esame, della quale si può valutare il comportamento o le caratteristiche, intrinseche o che mostra all'esterno, in diversi stadi del suo ciclo di vita;
 - ◆ l'utente che di questa entità fa uso per i propri fini, di quale si può rilevare la soddisfazione, la capacità di farne uso.
- Quando la fonte è l'utente si parla di valutazione della qualità "percepita" e bisogna considerare l'influenza di fattori soggettivi che potrebbero in parte condizionare la valutazione.

Tecniche di valutazione

- ◆ Le tecniche di valutazione devono permettere di coprire tutte le componenti di un prodotto (codice sorgente, documentazione d'uso, operativa, di manutenzione, dati, elementi di processo che influenzano la qualità del prodotto, effetti dell'uso del prodotto, qualità in uso).
- ◆ La scelta della tecnica dipende dallo stadio di lavorazione del prodotto e dagli obiettivi della valutazione (valutazione della qualità interna, esterna o in uso e da quale caratteristica si vuole valutare).
- ◆ A seconda del caso saranno appropriate tecniche di test statico (glass box), dinamico/prestazionale, field test di usabilità, test black box funzionali etc..

Metriche per la valutazione

Per ogni tipo di valutazione, bisogna rifarsi alle metriche opportune, a seconda di una valutazione diretta a caratteristiche interne, esterne od in uso del prodotto, come specificato nel modello 9126.



Elaborazioni dei dati rilevati

- Di norma, una singola misura non permette di esprimere immediatamente una valutazione su una caratteristica o sottocaratteristica di qualità.
- ◆ Dire che il tempo di risposta di una applicazione è "1 secondo" può essere soddisfacente in certi casi ed inaccettabile in altri, dipende dal contesto.
- ◆ Generalmente, per facilitare la valutazione del livello di possesso delle caratteristiche di qualità, le misure rilevate vanno trasformate, utilizzando una opportuna funzione, in una misura "derivata", che può essere definita come il "livello di possesso di una determinata caratteristica" (processo di *rating*).

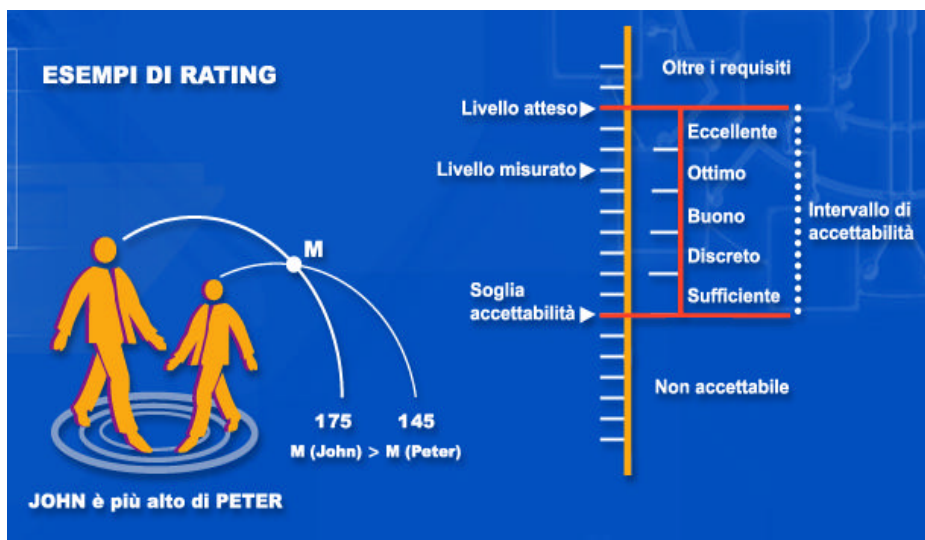
Criteri di rating

- Vi sono diversi possibili criteri di rating descritti in letteratura, ma nessuno è universalmente accettato per il settore dei prodotti informatici.
 - ◆ La scelta della corrispondenza tra misure e giudizio è fortemente soggettiva, e va fatta caso per caso secondo i dati di contesto.
 - ◆ Una *overview* del metodo da seguire per mappare misure quantitative su una scala qualitativa è nella norma ISO/IEC 14598-1.

Come effettuare il rating delle misure

- Un possibile procedimento di rating:
 - ◆ il valore "minimo" della scala associata alla misura derivata può essere fatto corrispondere con il valore più basso rilevabile della misura primaria,
 - ◆ il valore massimo è posto in corrispondenza al valore massimo misurabile per la misura primaria,
 - ◆ utilizzando una scala ad intervallo, l'intervallo dei valori misurabili della misura primaria va distribuito uniformemente nell'intervallo definito dal codominio della misura derivata,
 - ◆ a valori crescenti delle misure primarie deve corrispondere un *proporzionale* maggiore valore delle misure derivate.

Esempi di rating



Pericoli del “cattivo” uso del rating

- La traduzione di una scala a rapporti od intervalli (quelle di molte metriche 9126) in una scala ordinale (quella utilizzata per esprimere il giudizio) non ha fondamento "scientifico".
 - ◆ Una trasposizione tra scale differenti è corretta quando in ambedue i riferimenti la distanza tra due valori ha lo stesso significato fisico. Ad esempio, “A è alto il doppio di B solo quando A è alto 180cm e B 90cm”.
 - ◆ Ciò non è quasi mai vero nel caso di una valutazione della qualità su scala ordinale. E' difficile dimostrare, ad esempio, che la usabilità di un prodotto è esattamente il doppio, o un terzo, od una frazione, di quella di un altro.

Il rating delle misure - Manutenibilità

Ad es. per la manutenibilità, si può avere, utilizzando come criterio il livello di strutturazione del codice e come metrica il numero di McCabe, il rating seguente:

Complessità ciclomatica	Valutazione dei rischi di modifica	Manutenibilità
1-10	Un programma senza rischi di modifica	Alta
11-20	Rischi moderati	Buona
21-50	Alti rischi	Bassa
>50	Un programma non testabile, rischi molto alti	Insufficiente

↑
Scala ad intervalli

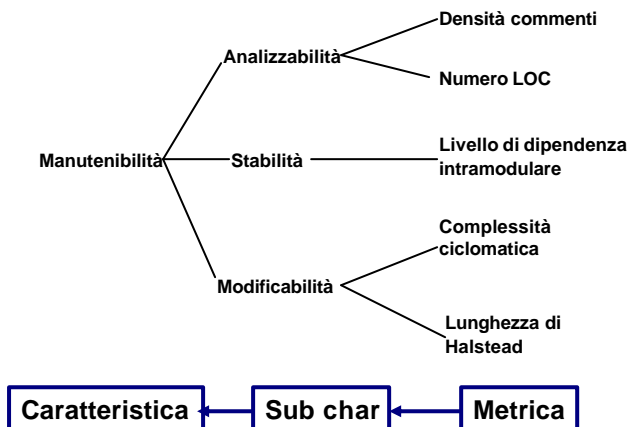
↑
Scala ordinale

Alberi di valutazione - Principi

- Poiché in genere una caratteristica è misurata con più metriche, per determinare quanto contribuisce ogni singola metrica alla valutazione della caratteristica si può utilizzare un albero di valutazione, pesato.
 - ◆ Nell'albero, caratteristiche (e sottocaratteristiche), e metriche, sono i nodi dell'albero (le metriche sono le foglie, o nodi terminali). La pesatura assegna ad ogni nodo un valore nell'intervallo $[0,1]$ dell'insieme dei numeri reali, che riflette l'importanza relativa del singolo nodo ai fini del determinare il valore della radice. La somma dei valori relativi assegnati a tutti i nodi con un padre comune deve essere uguale ad 1.

Alberi di valutazione - esempio (1)

Esempio per la valutazione della manutenibilità.



Alberi di valutazione - esempio (2)

Esempio per la valutazione della affidabilità.

Caratteristica	Sotto-carat- teso	peso	indicatore	peso	Val. ass.	metrica	Peso	Val. ass.
AFFIDABILITÀ	Maturità	0,2	densità di fermi in un periodo	1	0,2	n° di errori nel periodo	0,7	0,14
						n° di errori / tempo operativo	0,3	0,06
	Tolleranza	0,3	Capacità di evitare fermi in caso di errori	0,7	0,21	Numero di fermi / numero degli inconvenienti	1	0,21
						Capacità di ripristino	0,3	0,09
	Recoverabilità	0,5	Disponibilità	0,4	0,2	Tempo operativo effettivo / tempo operativo teorico	1	0,2
						Tempo di ripartenza dopo un errore	0,6	0,3

Valori di soglia - Principi

- Il valore di soglia è il valore atteso per le misure rilevabili, in riferimento al quale sono giudicate le misure rilevate durante le verifiche. Le soglie possono essere espresse in diversi modi:
 - ◆ come *valore medio* accettabile delle misure rilevate,
 - ◆ come *valore assoluto* (in alto od in basso) che delimita la accettabilità del prodotto,
 - ◆ come *distribuzione di frequenza* delle misure (ad es. il 90% deve rientrare in un dato intervallo, il 7% in un altro, il 3% in un altro ancora),
 - ◆ come *numero* di misure accettabile che può non rientrare in un dato intervallo di valori.

Definire i valori di soglia

- ◆ Per fissare le soglie, va considerato che aumentando la qualità richiesta aumenta anche il costo di produzione del software.
- ◆ Inoltre, superata una certa soglia, il rapporto qualità-costi di produzione non è lineare: a piccoli incrementi di qualità attesa corrispondono grandi aumenti dei costi di produzione.

Il piano di valutazione - indice

- Nel piano di valutazione vanno identificati:
 - ◆ Finalità della valutazione (obiettivi)
 - ◆ Elementi da valutare
 - ◆ Caratteristiche indagate ed obiettivi quantitativi
 - ◆ *Schedule* delle verifiche: tempo e risorse allocate
 - ◆ Responsabilità nell'effettuare le verifiche
 - ◆ Misure da rilevare (e metriche da utilizzare)
 - ◆ Criteri di elaborazione delle misure
 - ◆ Procedure di registrazione dei risultati
 - ◆ Ambiente hardware e software per il test
 - ◆ Casi di prova utilizzati

Effettuare la valutazione

- Scomponibile logicamente in tre passi:
 - ◆ nella *misurazione* sono rilevate le misure selezionate nella fase di pianificazione della valutazione;
 - ◆ nel *rating* sono mappati i valori rilevati su una scala di valutazione, utilizzando dei criteri;
 - ◆ nell'*assessment* si confrontano gli obiettivi che si vogliono raggiungere con lo stato attuale delle misure, determinando l'eventuale gap esistente.

Relazioni 9126-14598

