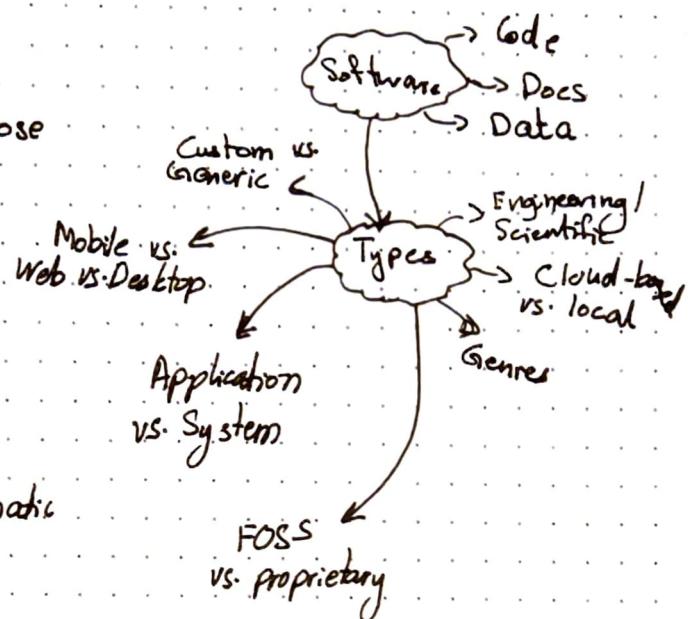


"Garbage in, garbage out."

Quality → fitness for purpose



Engineering: finding efficient solutions for real-life problems in a systematic and efficient manner.

Problems compound. (snowball effect, domino effect).

ENIAC → Softwares → First CS dept. → Software Engineering
1940s late 1940s # in US 1962

THRAC-25 mid-late 1980s Denver, Colorado Airport catastrophe Washington prison malfunction: over 3000 prisoners released earlier

Toyota recalled 600,000 'Prius' cars 2016

Nissan recalled ~3.5m Flex

WHY SQE is a separate field?

- nature (virtual/logical vs. physical)
- size
- team setting
- formalization
- interfaces
- personnel turnover

Algorithmic
Morality

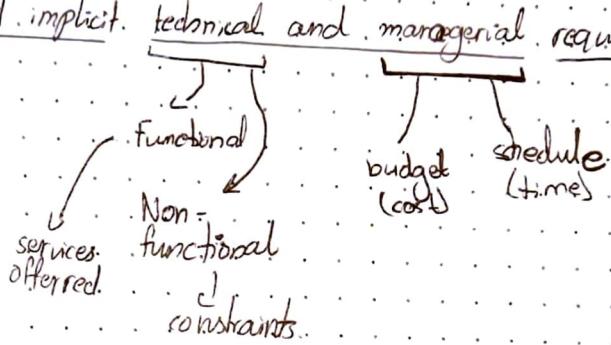
Assurance Software Quality Engineering.

Quality:

fitness for purpose

- conformance to explicit and implicit technical and managerial requirements

Perception of Quality ≠ Quality



Software Quality Problems:

- | | |
|--------------------------|---|
| <u>Software Supply</u> | <ul style="list-style-type: none"> - <u>Error</u>: 'to err is human' ... incorrect action or inaction |
| <u>Software Consumer</u> | <ul style="list-style-type: none"> - <u>Fault</u>: a state of the software ... not all errors become faults. |
| | <ul style="list-style-type: none"> - <u>Failure</u>: an event ... not all faults become failures. |
| | <ul style="list-style-type: none"> - <u>Incident</u>: detected failure ... not all failures get noticed |

↳ unstructured activity, process oriented

Software Quality Assurance: the set of systematic, planned activities required to provide adequate confidence in that the software produced through the software development process conforms to the explicit and implicit, technical and managerial requirements.

↳ subset of SQA, product oriented

Software Quality Control

"A defect is a delivered fault."

- Bessman not for us

"An accident is a severe failure"

Capers Jones → Software Project Management

"Short-cutting a day of QA activity early in the project is likely to add 3 to 10 days of unnecessary activity downstream."

- Capers Jones

- The most important deliverable of Req. Engineering:



Quality Factors → dimensions, attributes, content groups, SRS headings
ilities

(classic, 1977) → An abstraction of
McCall's Quality Factor Model → the real thing:
day-to-day operations → total 11 Quality Factors
product operation (5) → product change (3) → product revision (3) → product transition (3)
outside-world interaction

- | | | |
|---------------|-------------------|--------------------|
| - Correctness | - Maintainability | - Portability |
| - Reliability | - Flexibility | - Reusability |
| - Integrity | - Testability | - Interoperability |
| - Efficiency | | |
| - Usability | | |

Correctness: functional requirements, output-mission, output accuracy, output completeness, output up-to-date-ness, availability (response time), standards

Reliability: failure rates, probabilities

critical systems

Integrity: security (Authentication & Authorization)

Efficiency: optimal use of hardware (CPU, RAM, storage, communication bandwidth, battery life)

Usability: (ease-of-use)

- operations

- training

Question 03 (b)

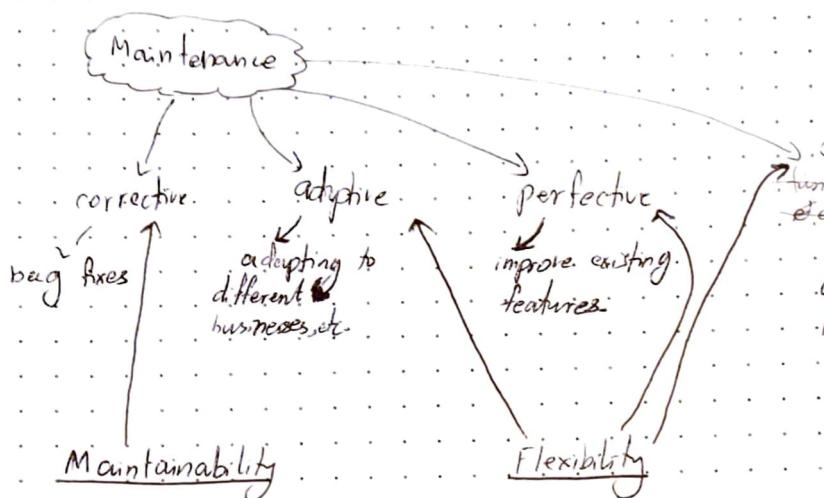
SQE LOS 04-09-2024

BIOS does POST

Power-On Self Test

McCall's model

Product Review: change and maintenance



Software supplier SRSs → Software Client/Consumer SRS sc.

Readability certain factors unique to each SRS version.

additive functionality enhancing add. features/ functionality

Testability

- easy to test
- intermediate results or logs
- allow operators and maintainers to test and diagnose the software
- both comments and readability are important
- comments are not an excuse for bad programming practices

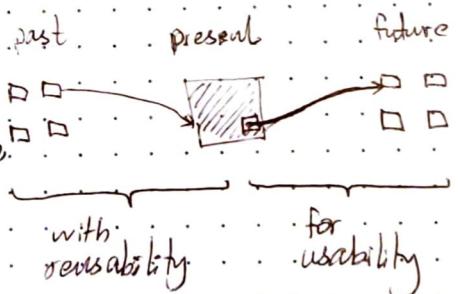
Product Transition:

- Portability: cross-platform support.

- Reusability
- for
- with

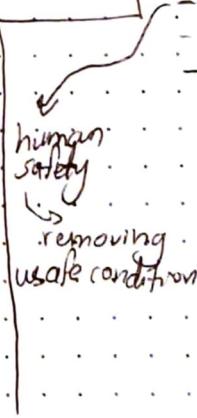
- improve documentation
- increase cohesion
- decrease coupling

- Interoperability



Alternative QF Models

- Evans & Marciniak (1987)
- Deutsch & Willis (1988)
- Verifiability → mathematical specifications
- Expandability ≈ flexibility
- Safety
- Survivability ≈ Reliability
- Manageability



- standard output formats

→ administrative tools for change management

Project Managers, UI/UX, programmers, testers, maintainers, devops, etc.
↳ many people can make mistakes.

Components of SQA system

Pre-Project Components

- Contract Review

- RFP → Request for Proposal

- Draft → Draft Review

SC evaluates submissions

↳ one gets a contract.

contract draft → review

- development & quality plans

- cost, resource estimations

- WBS, Risk plans

Life-cycle components

- Reviews

- team-based meetings

- variable degree of formality

- types: formal technical review (most formal),

(↳ FTR, Design Reviews, FDR)

Done by senior people

- Testing

- regression, smoke, automated, manual, unit, integration, other tests
black box, white box

- Expert Opinions

- outside reviews, etc.

- Third Party SUs,

SS - SC
SUS

Applicable
to projects at
hand

Applicable
organization-wide

QA applies to maintenance
as well

Component Categories

→ Pre-project

→ Life-cycle

→ Infrastructure

→ Management

→ Standards

→ Human

SS SC

↳ contract

Problem
Analysis
Phase

Solution
Phase

Inc. BE A D J T D/O O-M DR

Part 1

Part 2

Peer Reviews

↳ FTR, Design Reviews, FDR
(least formal)

COTS
/ ROTS, GOTS

Commercial

off-the-shelf

components.

Question 03(b)

SQE : L06 (continued)

Infrastructure Components

- Procedures & Work Instructions
- Supporting Quality Devices
 - Templates (SRS, project plan, risk plan, etc.)

- Training & Certifications
 - Training improves

PMP
PMI
ACP

- Software Configuration Management
 - Version control
 - Git, subversion, etc.

Scrum Master

- Analysis of Past-Projects
 - What went wrong
 - Where most mistakes occur

Management Components

- Progress Control

Managers do planning, monitoring, controlling

- Quality Metrics

- defect density

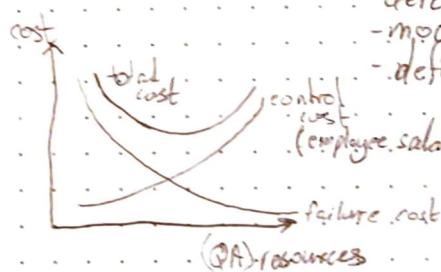
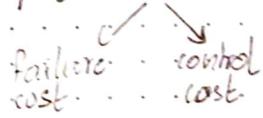
- productivity

- defect severity → weighted defects

↳ module size

- Cost

- minimize project cost



M1	M2
100	500

Quality dependence:

- defect severity
- module size
- defect density

Standards Components

- Quality Management (What)

- CMMI: Capability Maturity Model Integrated

↳ developed by CMU

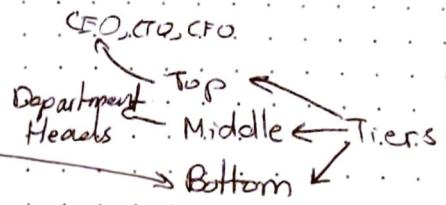
- ISO 90003

- Project Process (How)

- IEEE 1212

- ISO/IEEE/IEC 12207

Human Components
→ Employees, team members, project managers.



Factors Affecting QA system

- organization size \propto QA effort required
- project complexity \propto QA effort required
- experience (relevant to the project domain) \propto QA effort required
- 3rd party engagement \propto QA effort required
- acquaintance of team members with each other \propto QA effort required
- extent of reuse (with or reuse) \propto QA effort required
- qualification \propto QA effort required

Question 03(b)

SQE LOT 108 11-09-2024
13-09-2024

BRING CALCULATOR (esp unique selling point)

INTEGRATION OF SQA ACTIVITIES IN SDLC

Usability of drop-down
recognition vs. Recall

Process Control vs.
Quality Management:
What vs. How,
not product vs.
process oriented

Software Processes

Process: a set of activities approached in a specific order.

↳ approach, roadmap,

Process Model: an abstraction of the process.

Spiral is a meta-process.

change-friendliness

Conventional,
Rigorous, Heavy-weight

Traditional

I Know It When
I See It
Syndrome

light-weight,
lean

Agile

User Acceptance
→ Testing (UAT)
→ Requirements
→ OOP
→ preset velocity

XP → Planning
→ IP → Gathering
→ CRC → User Stories
→ no. of user stories implemented in the current release

Scrum → Commitment
what user stories by what deadline
KISS ← Minimalist Design
design prototypes called spike solutions

Large Projects
called
solutions

Small Projects
or medium

Naturally Modularizable Problem

Rapid Delivery Required

- Kent Beck's book on XP:
- Refactoring by Martin Fowler

Evolutionary

Throw-away prototypes
Build-upon prototypes

Kiwiwi Prototyping

Planning Problem: when will it end

Spiral

Planning Problem: when will it end

Waterfall (classical)

Heavy-weight process

Non-Evolutionary

Incremental (Pure)

no. of features

RAD

↳ Rapid Time, ~60-90 days

large projects, stable reqs, inacceptable when: unscrupulous customer, volatile reqs, feedback needed

only in hybrid versions

Change welcome

High Expertise

User Stories

As a who, I want what so that why

- Priorities (SE) ↗
Cost (SS) ↘ value

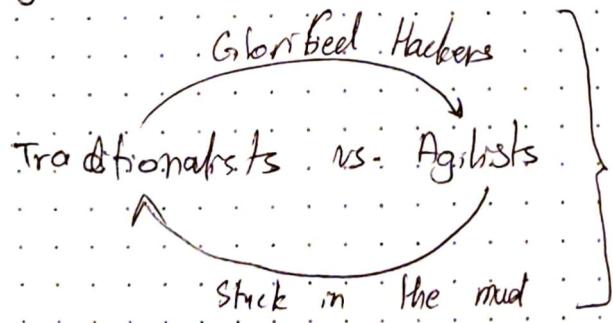
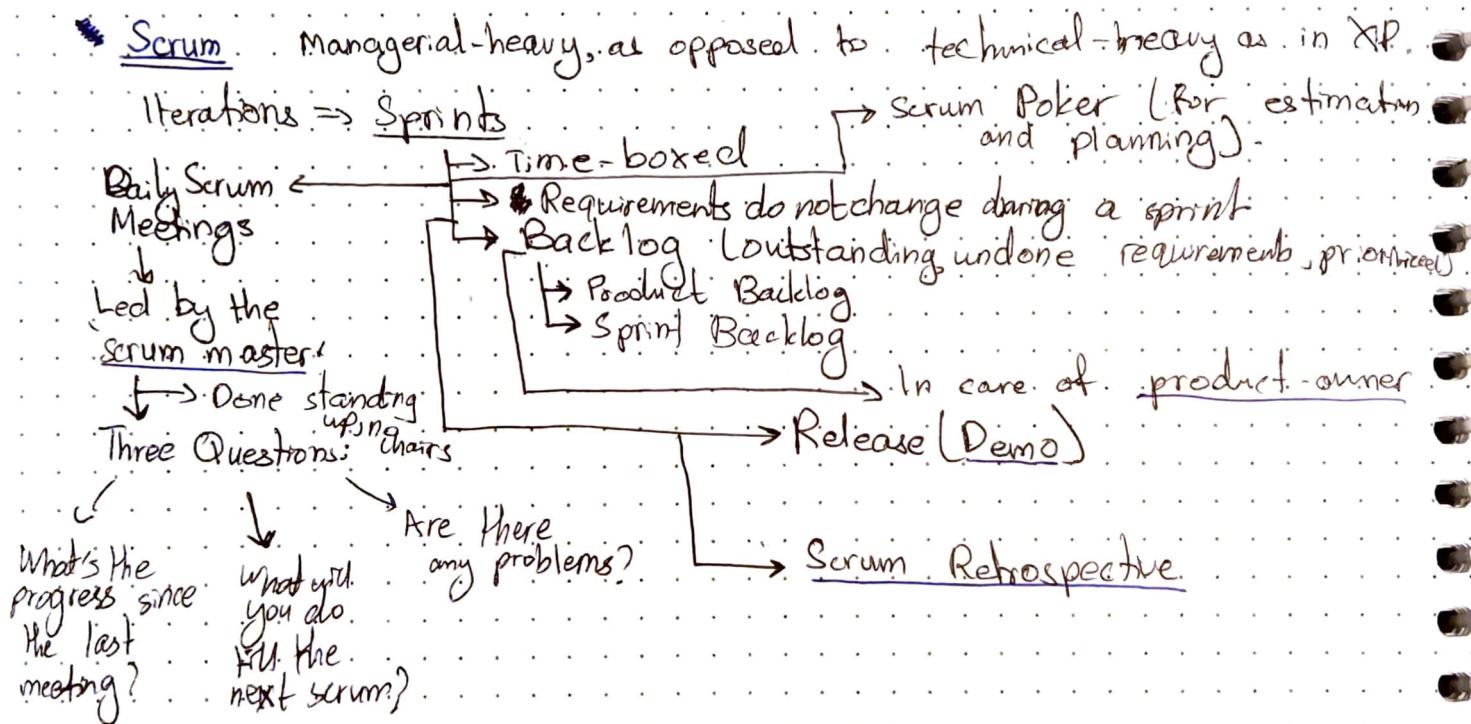
February 2001, 17 people signed the agile manifesto in Utah:

Kent Beck, Robert C. Martin

→ later extended to other fields like manufacturing, design, and project management.

Agile Manifesto

1. Individuals & Interactions vs. Processes & tools
2. Working software vs. Comprehensive Documentation
3. Customer communication vs. Contract Negotiation
4. Responding to change vs. Following a plan



Barry Boehm \rightarrow Balancing Agility and Discipline

Dimensions of SQA

Defect Removal Effectiveness & Cost Model

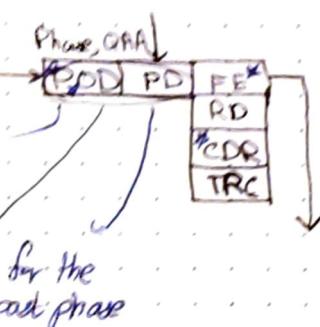
- QA activities are filters, not 100% accurate, but still cost something.

- Defect

- Filtering Effectiveness

- Cost Data in Relative Terms

Design Review



15,000
30.1.

Coding Inspection		
300	100	60
240		
*	16cu	
		160
		3,340

* is past data

for the current phase

for the past phase

Total Defects

$$\text{reference points} \leftarrow \text{RE} = 1\text{cu}$$

$$\text{coding} = 16\text{cu}$$

$$RD = (PDD + PD) \times \frac{FE}{100}$$

$$TRC = RD \times CDR$$

Assumptions:

- All SQA activities act like filters
- Past data is accurate and still relevant.

Verification: are we making the product right?

Validation: are we making the right product?

Qualification:

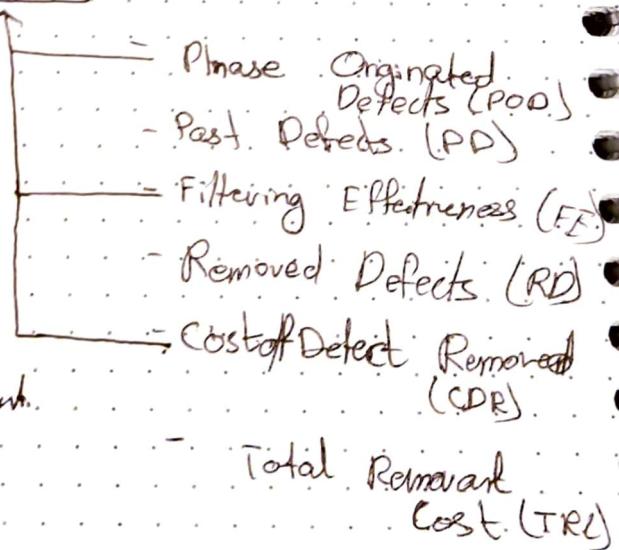
- coding conventions
- good SE practices
- general standards

Circle Area Examples

- Verification: Is it programmed as it is?

- Validation: Did the customer actually need this?

Past Data



Question 03(b)

SQE (09)

No.	Software Development Phase	Engineering defects
1	Requirements	35/h
2	Design	40/h
3	Coding (30% coding, 10% integration)	
4	Documentation	10/h

Standard QA plan (Plan A)

No.	QA activity	Finding	Cost of finding defects
1	Requirement Specification review	50%	1 cu
2	Design review	50%	2.5 cu
3	Unit Test - Code	50%	6.5 cu
4	Integration test	50%	16 cu
5	Documentation review	50%	16 cu
6	System test	50%	40 cu
7	Operations phase	100%	110 cu

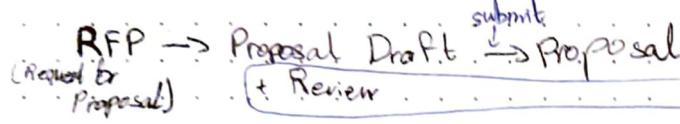
Comprehensive SQA Plan

No.	QA Activity	Finding	Cost of finding defects
1	Requirement specification review	60%	1
2	Design inspection	70%	2.5
3	Design review	60%	2.5
4	Code inspection	70%	6.5
5	Code + unit unit test	90%	6.5
6	Integration test	60%	16
7	Documentation review	60%	16
8	System test	60%	40
9	Operations phase	100%	110

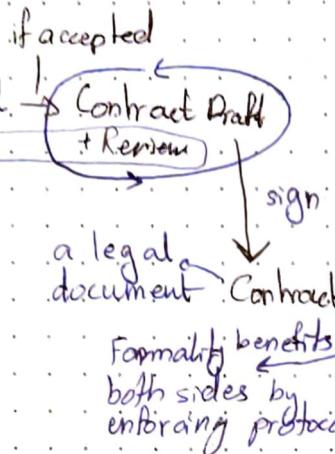
Contract Review

Pre-project component

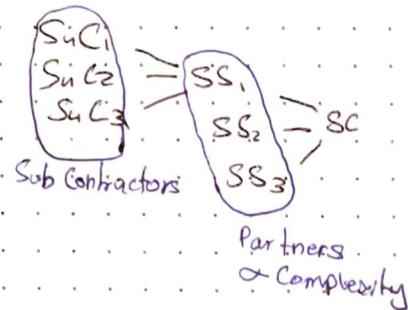
Crime has consequences



- Requirements clarified
 - ↳ scope
- Alternative Approaches
- Relationship Formality
 - focal person representative
 - Project deliverables
 - Change approval process
- Risks
 - technological
 - non-technical (personnel turnover, economic, political conditions)
- Estimates
 - Effort
 - Duration
- Capacity / Capability
 - Can the supplier do it?
 - Can the consumer pay us?
- Participants
- Intellectual Property
 - Who gets to own the copyrights.



↳ SC & A activities should be scheduled.



Goals of Contract Draft Reviews

- no unclarified issues
- understandings documented
- no "new" changes

no ambiguity
no less
no more

extent of the contract draft review

- depends on project size, complexity, organisational complexity, acquaintance/familiarity/expertise

characteristics of Bad Contracts

- unrealistic estimates
- loose requirements

estimates can vary based on the software development process. But the choice of the process need not be documented

FORMAL TECHNICAL REVIEWS → lifecycle components.

Review: critical evaluation/appraisal; formal

- Used in thesis reviews, FYP reviews
- double-blind review: no one knows who is the other person
- Reviews are used in other fields

- Reviews also focus on code, but primarily on docs.

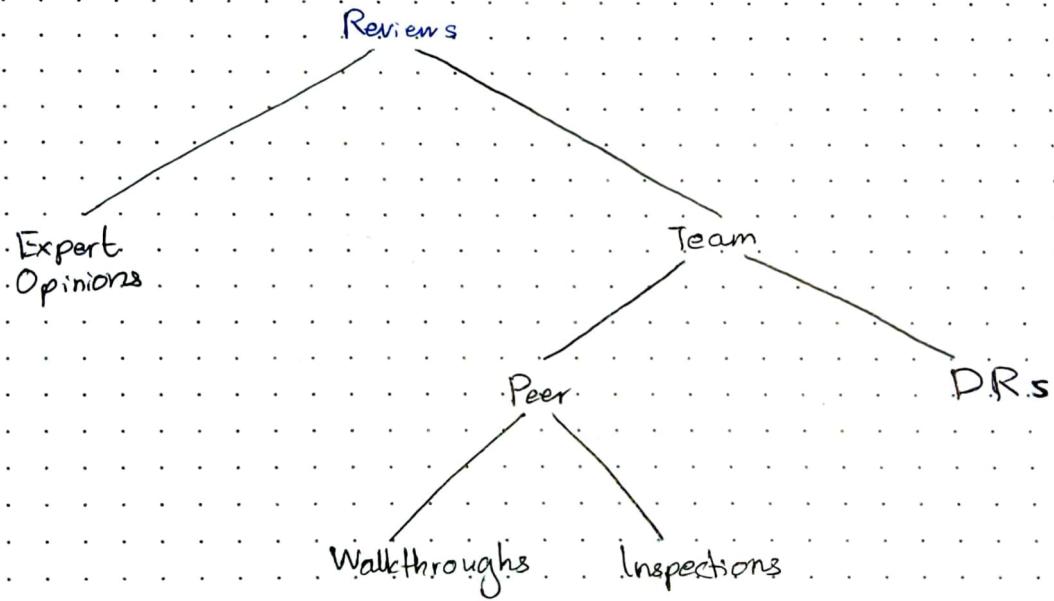
- Reviews are generally good at detecting defects. They act as filters to remove defects.

Primary focus of:

- review
- Testing

Software =

- Programs
- Docs
- Data



Emphasis on objectives varies from review to review

Direct Objectives (deal with the project at hand ???)

- detect and remove as many upstream faults (check incomplete, missing info/parts)
- identify new risks
- older risks already identified in risk plan
- detect deviations from templates/conventional standards
 - deviation cause problems downstream in & maintenance
- grant or op. deny approval

Indirect Objectives (deal with all projects)

- exchange professional knowledge (in meetings of professionals)
- record/analyze data related to faults

Expert Opinions

Outside consultants,
senior people

Design Reviews (a.k.a. Formal Design Reviews) (a.k.a. Formal Technical Reviews)

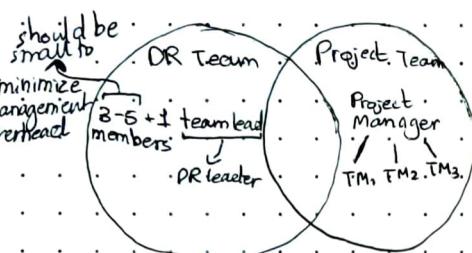
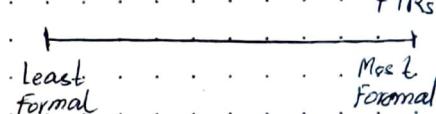
↳ most formal, most authoritative

4 segments of DR: ~~before~~ planning, before, during, after.

When outsiders are needed?

- when the SC is small and doesn't have enough manpower
- when the in-house expertise is not enough
- resolve disagreement

The formality spectrum



People may overlap, but play different roles.

Planning: assemble the team and appoint DR Leader.

DR Leader:

- should not be the project manager of the project at hand
- should be preferably more senior than the project manager
- may be the project manager of another project
- should be familiar with the type of project

DR team members:

- majority of these should not be from the project team
- ~~only~~ senior people could join in a minority
- should be diverse (to get a diverse perspective of feedback)

Before

- DR leader assigns responsibilities and sets a schedule
- small projects may be reviewed in entirety by everyone, otherwise work will be divided
- DR is reviewed against documents. documents against checklists
- DR team comes prepared with comments
- Project team prepares short presentation

During

- DR session starts with a short presentation
- The comments and action items are discussed along with assignees, due dates, and verifications
- DR leader is usually the scribe
- Decision (full/partial approval, or denial) is ~~not~~ made
- As many sessions may be held as needed, but one session should not last more than 2 hours.

DR Leader ensures:

- no personal criticisms
- schedule is followed
- offline discussion of internal conflicts
- external experts may be involved

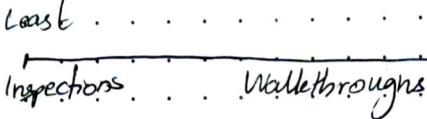
After

- PR leader prepares a DR report quickly and shared with stakeholders
- Follow-ups to reflect on the ~~PR~~ PR
- If something is not fixed satisfactorily, another follow-up is needed.

it is

PEER Reviews 27-09-2024

Formality Least



Most

DRs

Team composition:

- Peers make up a PR team
↳ peers of the author of the document

Inspections

Walkthroughs → Quicker but less comprehensive

Planning Phase

Leader: Moderator

Team members:

- Author(s)
- a designer
- a programmer
- a tester
- Project Manager should not be the leader

Leader: Coordinator

Team Members:

- Author(s)
- a standards enforcer
- a user representative
- a maintenance person
- Project Manager may be the leader

Before Phase

- Overview Meeting

- Moderator should give an overview of the document being ~~inspected~~ if team members are outsiders.

- Thorough Preparation:

- Discuss Comments
- Evaluate documents against checklists

- Brief Preparation

- Skimming the documents,
- possibly giving ~~some~~ comments

During Phase

- Short presentation

- Programmer presents the document

- Sessions like in DRs

- No decisions taken

- Short presentation

- document is presented by the authors

- Sessions like in DRs

- No decisions taken

After Phase

- 2 reports → findings
- Follow-up

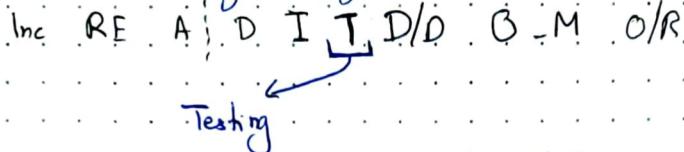
- 1 report → findings

- no follow-up

Effort vs. Duration
- person hours
- days
- weeks
- months
- years

SQE 02-10-2024

Software Testing Strategies



- the classical, oldest, last-line-of-defense against faults.

- takes the most amount of SQA resources.

- Testing requires execution of code

- not every execution of code ~~means~~ is testing

- Testing requires a certain mindset that faults need to be uncovered.

→ high-level

→ low-level

Software	Testing	Strategies	Techniques
- program/code	- execution	- what	- how
- data	- intention to find faults	- who	
- documentation	- formal - approval - third-party	- when	

Designers:

- use-cases first w.r.t GUI

- modularity, coupling, cohesion

- functional independence

Strategy vs. Technique

Objectives of Testing

Direct:

- identify as many faults as possible and fix them

- do this efficiently

- acceptable quality

Indirect:

- to compile a record of fault data

Software Testing Strategies

General Testing

Unit Testing

~~test integrated related units~~

- test a single unit

- unit = set of related classes/functions

- data structures - UI

- algorithms

- loops

- conditions

- big-bang approach

- incremental approach

- Top-down

- Bottom-up

- Depth first
- Breadth first

Integration Testing

- test integrated modules

Validation Testing

- test conformance to requirements

System Testing

- non-functional requirements

- performance test intended

- stress test on ~~related~~ hardware

- more related to embedded

systems.

Coders:

- programmable or moonshot?

Testers:

- quantifiable

- reqs. testable?

Designers:

- use-cases first w.r.t GUI

- modularity, coupling, cohesion

- functional independence

Integration TestingBig-BangIncremental

Integrate everything together at the same time.

Works well for small projects.

Pinpointing the errors in the modules is more difficult.

The Hamster EffectDecision-Makers vs. Work horses

Requires "Dummy" Modules.

Integrate step-by-step.

Which one to use?
Follow the development approach.

Overhead Cost.

Top-Down ("Stubs") Bottom-Up ("Drivers")

Stage 3

User Acceptance Testing (UAT)

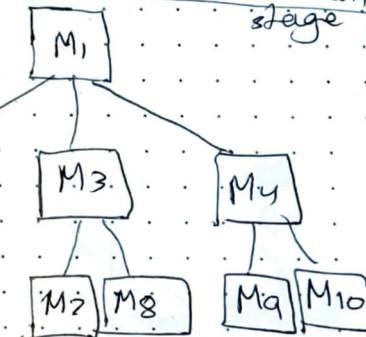
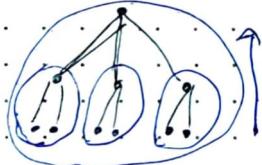
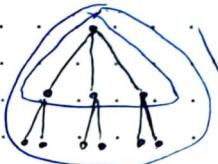
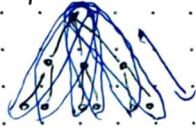
- Bespoke/custom products require UAT.
- Done in natural environment

"Details", "Error Message" are ambiguous.

- Database backend things should not be present in the SRS.

- Formatting issue should be mentioned.

Develop and unit tested in each stage

Bottom-Up Breadth-FirstBottom-Up Depth-FirstTop-Down Breadth-FirstTop-Down Depth-FirstValidation Testing:

- f. end-users involved; not professional testers.

alpha testing

~~Hamster Effect~~ - end-users test the software in a controlled environment.

beta testing

- end-users test the software in their natural environment.
used for generic open-market products.

Smoke Testing:

- Electronics ~~short-circuit~~ ~~only~~ release smoke only when there's a short-circuit.

- tests whether

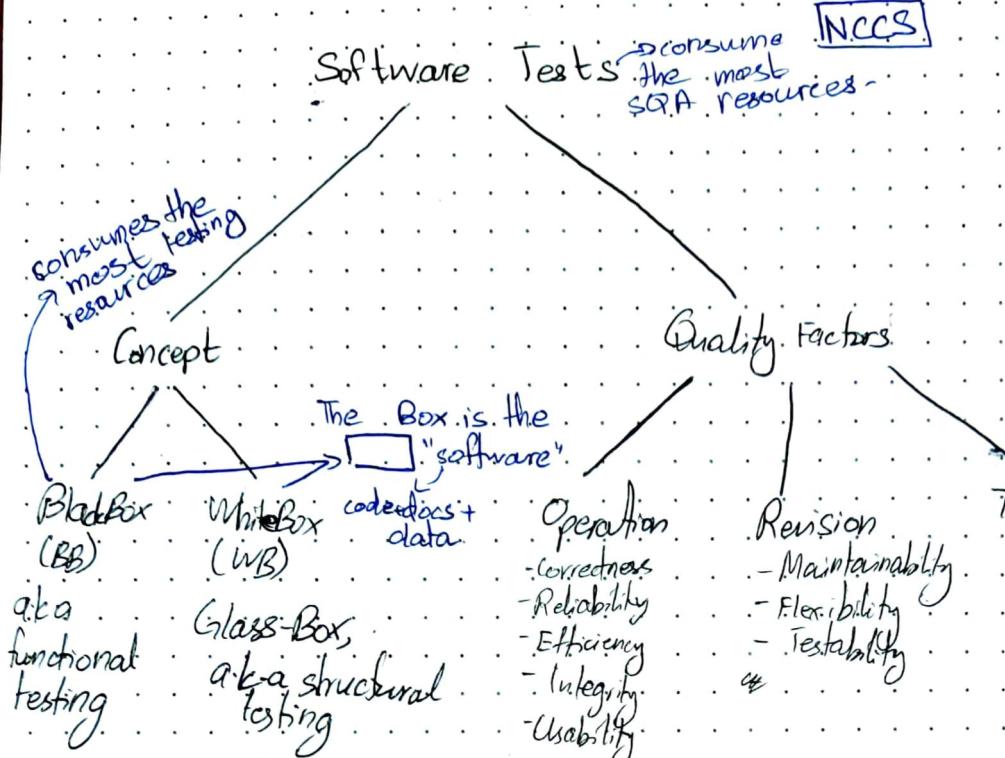
- software runs
- does not crash

- major features work (no show-stoppers)

Build-Verification TestingRegression Testing:

- Done after each change

- Test ~~subset~~ previous test cases to check if the change hampers previously working code.



Hybrid Versions.
Like Grey-Boxtesting.

- Reliability (BB)
 - MTBF between failures
 - MTTR to repair
 - MDTM down-time per month.
 - Statistical models exist for reliability testing
- Efficiency (BB)
 - Stress tests
 - load testing + ~~stress~~
 - behavior / performance of the system under abnormal load.
 - durability testing
 - expose system to abnormal physical conditions.
- Integrity / Security Testing (monthly)
 - Malware, viruses, trojans
 - SQL injection
 - ethical hacking
- Usability Testing (mostly BB)
 - difficulties increase error chances
 - Training usability testing
 - Operational usability testing

- Correctness: WB / BB

- correctness of does (like manuals, user installation manuals, etc.)
- correctness of outputs
- ~~top~~ with/without focusing on details
- availability testing

- usually done with load testing
- response time as a function of load
- usually automated

- Maintainability (BB / WB)

- corrective modular designs, functional

- Flexibility (BB / WB)

- independence, low coupling, high-cohesion

- Testability (WB)

- intermediate results

~~- in-house checks~~

- Portability (BB)

- The software works on other systems or not

- Reusability (WB)

- satisfies criteria for inclusion in the re-use library

~~- coupling, cohesion, functional independence~~

- Interoperability (BB)

- Working with other software interfaces

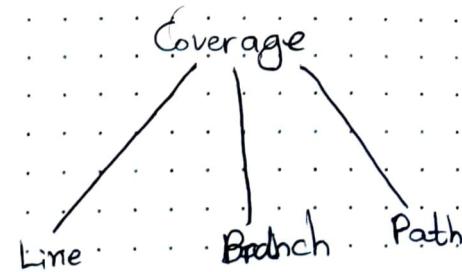
~~- systems~~

WHITE BOX TESTING

If there are 10 sequential, un-nested conditionals, there are $2^{10} = 1024$ possibilities.

EXHAUSTIVE TESTING IS NOT PRACTICAL.

- Selective testing is preferred unless the module is very critical.
 - All programming language statements must be executed at least once.
 - ↳ Line coverage = 100%
- 100% line coverage does not guarantee 100% branch coverage



This test case covers 100% of the lines, without covering all branch.

Each branch is

Basis Path Testing / Basic Path Testing

↳ Guarantees 100% Branch Coverage.

Code → Program Flow → Cyclomatic Complexity $V(G)$

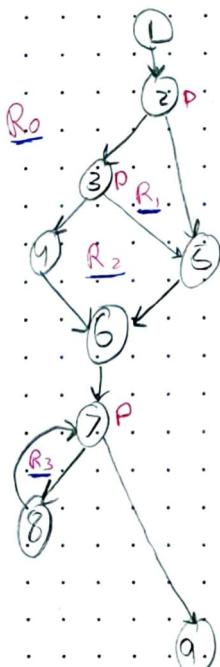
$$\text{Regions} \rightarrow \text{Enclosed Region}$$

$$R = ER + 1 = 3 + 1 = 4$$

$$V(G) \rightarrow P + 1 = 3 + 1 = 4$$

Edges
Nodes

Predicate Nodes: Nodes with 2 or more outgoing edges



Path → Branch → Line

→ has certain rules
→ Upper bound on the no. of basis paths

→ Basis Paths → Voila! 100% Branch coverage.

Lefty-truthy, righty-falsey.
Rules

1- Branch together all consecutive sequential non-branching statements in a single node.

2- All loop/if/else conditions become one node.

Upperbound on the amount of basis paths needed for 100% Branch coverage.

It is a UPPER BOUND if it is possible to achieve 100% branch coverage with lesser test-cases.

Set of Basis Paths

S#	Path
1.	1 → 2 → 5 → 6 → 7 → 9
2.	1 → 2 → 5 → 6 → 7 → 8 → 7 → 9
3.	1 → 2 → 3 → 4 → 6 → 7 → 9
4.	1 → 2 → 3 → 5 → 6 → 7 → 9

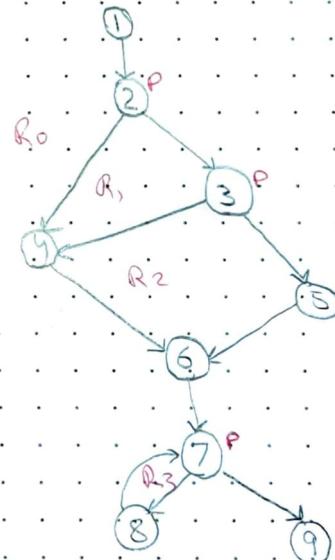
IDEs may calculate the cyclomatic complexity.
It can be done by identifying the nodes and creating an adjacency matrix for the graph.

$V(G)$ is used for test prioritization, nodes, and creating an adjacency matrix for the graph.
→ Independent (by definition) flow graph
→ Subsequent paths remain independent by adding at least one such edge which is not present in any of the previous/independent paths.

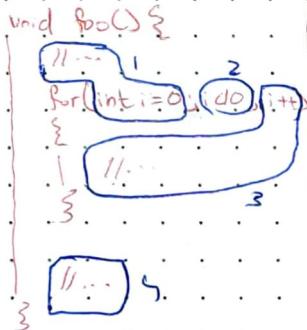
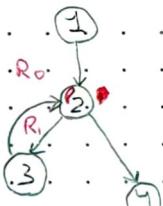
S#	Path
1.	1 → 2 → 3 → 4 → 6 → 7 → 8 → 7 → 9
2.	1 → 2 → 5 → 6 → 7 → 9
3.	1 → 2 → 3 → 5 → 6 → 7 → 9

$V(G)$ is the cyclomatic complexity, which is to be used for calculating the Weighted Methods per Class (WMC) in the CK set of O.O. Metrics.
 $WMC = \sum c_i$

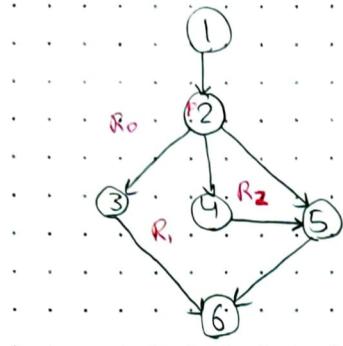
S#	Path
1.	1 → 2 → 3 → 5 → 6 → 7 → 8 → 7 → 9
2.	1 → 2 → 3 → 4 → 6 → 7 → 8 → 9
3.	1 → 2 → 3 → 5 → 6 → 7 → 9



S#	Path
1.	1 → 2 → 3 → 2 → 4

 $V(G)$

$$\begin{aligned} &= R_0 + FR + 1 = 1 + 1 = 2 \\ &= E - N + 2 = 4 - 4 + 2 = 2 \\ &= P + 1 = 1 + 1 = 2 \end{aligned}$$



$$\left. \begin{aligned} V(G) &= R = ER + 1 = 2 + 1 = 3 \\ &= P + 1 = 1 + 1 = 2 \\ &= E - N + 2 = 7 - 6 + 2 = 3 \end{aligned} \right\} \Rightarrow V(G) = 3$$

```

int selection(int option) {
    int value;
    int denominator;
    switch(option) {
        case 1:
            value = option + 1;
            denominator = 2;
            break;
        case 2:
            value = option - 1;
            denominator = 4;
            break;
        default:
            value = 0;
            denominator = 1;
            break;
    }
    return (value / denominator);
}
    
```

```

void foo() {
    ...
    bar();
    ...
}
    
```

function calls are non-branching.

```

void bar() {
    ...
}
    
```

- Recursive calls are just like function calls.

SQE L18 18-10-2024

Testing Condition Statements

- ↳ have more error chances
- ↳ if-else-if/switch
- ↳ use operators
 - relational $>$, $<$, \geq , \leq , $=$
 - logical $\&&$, $||$, $!$
 - incorrect operators
 - missing operators
 - parentheses

Boris Beizer: "Bugs lurk in corners and congregate at boundaries"

Nested loops:

- Begin with inner loop. keep all outer loops at min. iterations.
- Move out one loop at a time.

Concatenated:

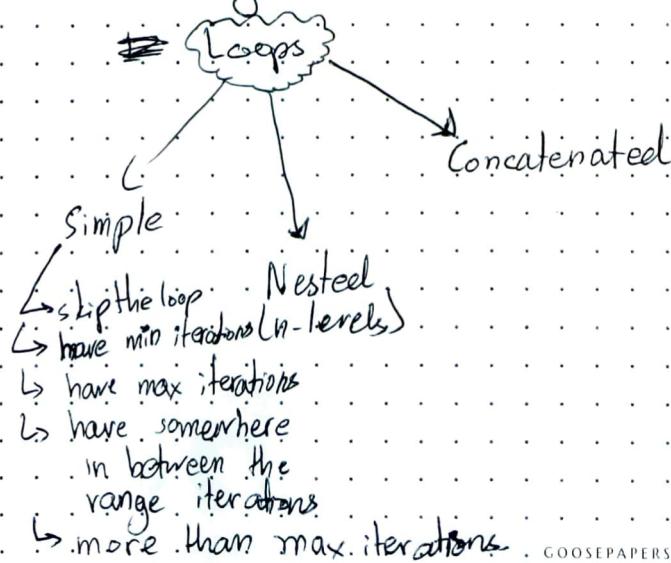
Independent scenario:

Do simple testing of each loop in sequence.

Dependent Scenario:

Treat loops as nested.

Loop Testing:



Advantages of White Box Testing

- Test Source code directly
- Coverage can be controlled directly
- Qualification Testing (conformance to standards) can also be tested.

Disadvantages:

- Resource-heavy, takes more time than black-box testing.
- Options are limited.

Chapter 6:Floyd's AlgorithmBlack-Box Correctness Testing

→ since source code is not available, we test the correctness of the outputs.

Equivalence Class Partitioning (ECP)

Input is partitioned into equivalence classes.

Equivalence Class: a set which contains inputs which make the software behave in the same manner.

- created for each variable, keeping in mind the software

CGPA

B1) 0.0 - 1.9 → 0.5, 1.2, etc.

B2) 2.0 - 4.0 → 2.1, 3.7, etc.

Invalid Equivalence Classes: sets of input values to which the software responds with specific error messages. Each type of invalid ~~eq~~ input will have a separate equivalence class.

Coverage in Correctness Testing:

We need minimum test-cases

Rules:

- 1- Write a separate test case for each invalid class and test only on e.c. in each test case.
- 2- Write a single test case for ~~all~~ the valid e.classes together.

The idea is that separately checking incorrect inputs easily identifies the underlying error.

→ Black-box test cases may be written even before the code is available. Although code is needed to execute test-cases, it is not needed to 'write' these test-cases.

Boundary Value Analysis (BVA)

The boundary values of all equivalence classes used for testing.

- can not be used for discrete variables, like gender, types, etc.

SC Example
↳ Swimming Centre

- Day of the week → Weekday → Week-end day
- Visitor's status → Member → One-time visitor
- Visit Time → Shift -1 → Shift -2
- Age → Young → Middle → Elderly

Possible Valid

$$\text{Combinations} = 2 \times 2 \times 2 \times 3 = 24$$

- Costs given in table 9.7, Galvin

Number of test-cases in 1A = maximum no. of equivalence classes of all input variable
 " in 1B = maximum no. of boundaries
 " in 2 = no. of invalid equivalence classes
 = twice the number of equivalence classes of a continuous input variable with the maximum no. of equivalence classes

$$\begin{array}{ccc}
 a=2 & \xrightarrow{\hspace{1cm}} & 2 \times 1 = 2 \\
 b=1 & \xrightarrow{\hspace{1cm}} & 2 / 1 = 2
 \end{array}
 \boxed{2 \times 1 = 2} \quad c=2$$

Advantages

- Cheaper than whitebox testing (where both can be used)
- Many options

no fine control on code coverage.
 Some control on conceptual coverage.

Disadvantages

- Qualification testing not possible
- No control on coverage
- We may get the correct output by chance

THE TESTING PROCESS

Decide
Methodology

- Acceptable level of quality
- Strategies & Techniques

Plan

- What do we test?
- Who performs the test?
SS/SC/consultants
- Where is the testing performed?
SS/SC/Consultants
- When do we stop testing? How much testing is enough?

Test Case
Design
(1-5)Damage
Level
A

(1-5)

Risk
Level
B

Reports are prepared
perhaps automated.

High-level decisions to be made at start of the project:

- Acceptable Level of Quality
- Strategies & Techniques
 - ↓ Incremental/Big-Bang
 - ↓ Top-down/bottom-up
 - ↓ Breadth-first/Depth-first

Stopping:

- stop when all is tested
 - assumes unlimited budget
 - may work for tiny projects
- stop when we have ~~run out of~~
money/time
 - surprisingly, very common.
- stop based on mathematical models
 - a threshold is defined for the "detect detection rate" based on past data
 - criticisms:
 - different severity-level defects treated equally
 - current data not reliable predictor of future data

- Fault seeding

- intentionally inject defects
- define threshold
- premise: correspondence between seeded and actual defects
- seeded faults removed before delivery

A → How dangerous is the damage?

B → How likely is the damage.

i - A+B

ii - kA + mB

Risk Exposure
(Expected damage)

Used in project management to prioritize project risks.

- dual independent testing teams..

TEST CASE DESIGN

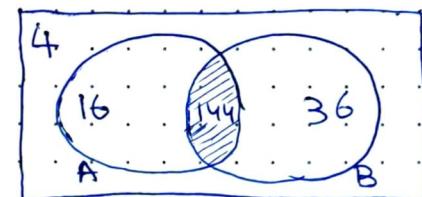
Test Case: a set of input values along with operating conditions and an expected output.

At least 1 test case for each course of action of every use-case.

Priority = "High", "Medium", "Low", ...

Dual Independent Testing Teams Method

N_a → no. of faults detected by team A
 N_b → no. of faults detected by team B
 N_{ab} → no. of faults detected by both teams
 N → total no. of faults
 T → threshold



Example:

$$\begin{aligned}
 N_a &= 160 \\
 N_b &= 180 \\
 N_{ab} &= 144 \\
 T &= 2.5\% = 0.025
 \end{aligned}$$

$$\begin{aligned}
 P(a)(b) &= \left(1 - \frac{N_a}{200}\right) \left(1 - \frac{N_b}{200}\right) \\
 &= \left(\frac{40}{200}\right) \left(\frac{20}{200}\right) \\
 &= 0.02 = 2.1\%
 \end{aligned}$$

since $2.1\% < 2.5\%$, testing may stop

$$\begin{aligned}
 N(a)(b) &= P(a)(b) \times N \\
 &= 0.02 \times 200 \\
 &= 4
 \end{aligned}$$

$$\begin{aligned}
 \frac{N_{ab}}{N} &= \frac{144}{200} = 0.72 \\
 P_a \times P_b &= \frac{160}{200} \times \frac{180}{200} = 0.72
 \end{aligned}$$

$$\begin{aligned}
 P(a)(b) &= \frac{\left(\frac{N_a N_b}{N} - N_a\right) \left(\frac{N_a N_b}{N} - N_b\right)}{\left(\frac{N_a N_b}{N}\right)^2} \\
 &= \frac{(N_a N_b - N_a^2)(N_a N_b - N_b^2)}{(N_a N_b)^2} \\
 &= \frac{(N_a N_b - N_a^2)(N_a N_b - N_b^2)}{(N_a N_b)^2}
 \end{aligned}$$

Using simplified formula:

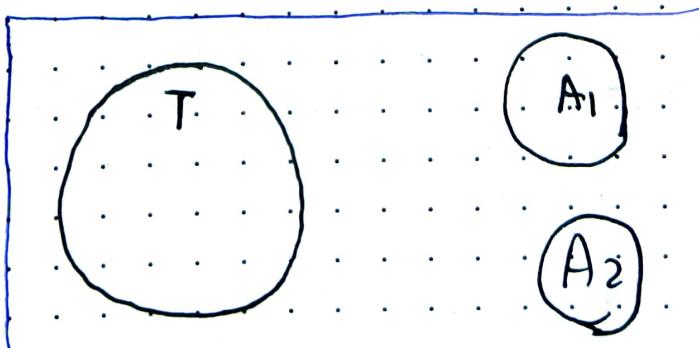
$$\begin{aligned}
 P(a)(b) &= (180 - 144)(160 - 144) \\
 &= 160 \times 180 \\
 &= 0.02
 \end{aligned}$$

Test Case Data Sources

Real life samples

Synthetic

- usually done when a manual process is automated.
- minimizes test-case preparation effort since inputs and outputs are already present.
- in case of pure random sampling.
- more execution effort since size of test-case file is large.
- test file is large because many sample data points are typical, and alternate cases are limited.
- done using ECP and BVA
- more effort in preparations since inputs and outputs need to be synthesized or calculated.
- less execution effort since size of test-case file is small.



Sample

Circles do not actually exist; these are "scenarios" or "strata".

These strata are defined from the standard and alternate courses of action of each test case.

Consensus: to add both synthetic and real-life data points to the test case file.

Rationale: unexpected things happen in real-life.

Automated Testing

- Code auditors
 - e.g. CheckStyle for Java
 - automate qualification testing
 - generates compliance/non-compliance/summary reports
 - requires definition of these standards as input
- Coverage Monitors
 - monitor line/branch/path/conceptual coverage
 - Jacoco (Java Code Coverage)
- Functional Testing
 - black-box functionality testing (black-box output)
 - Selenium, QTP, Cucumber, etc. (correctness tests)
- Load Testing
 - LoadRunner, Apache JMeter
 - ~~simulate~~ simulate high load scenarios (hits-per-minute, transactions-per-minute)
- Test Management
 - Bugzilla, Jira
- Usability Tools
 - Maze,

Automated

vs.

Manual

- require tool licensing, training costs.
- test-case preparation is still manual.
- overcome human shortcomings, like forgetfulness, boredom, demotivation.
- generate more accurate and complete bug reports.

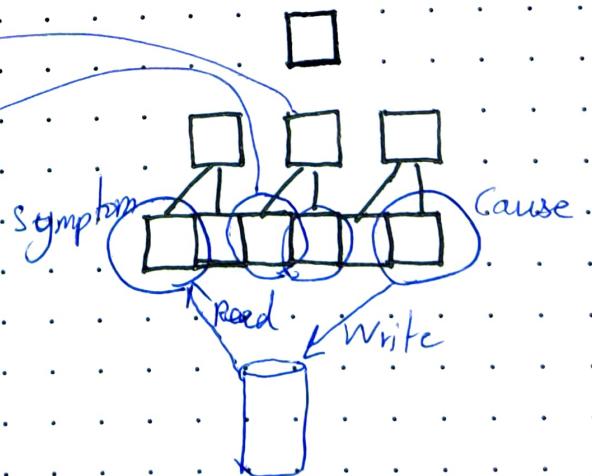
Not everything can be / has been automated in testing.

→ Trying to connect symptoms to causes.
Debugging IS NOT TESTING!!

Occurs as a consequence of a test-case that successfully fails.

This, not expecting rounded-up results can cause problems

These two may produce the same, correct output after rounding up, which is a non-problem.



Debugging may be intermittent: sometimes it gives correct output, other times not.

Debugging Strategies:

- brute force.
check everything when no suspected cause ~~problems~~ found.
- back tracking.
check symptom code, then the code calling the code, and so on.
- cause elimination.
write extra test cases to ensure that the suspected problem is the actual problem.

Object-Oriented Software Testing

Overall testing strategy:

small → large

unit → integration → validation → system testing

same for conventional and OO software.

In OO software, unit and integration testing change meanings.

OO Unit Testing

- Unit is a class (data and behavior).

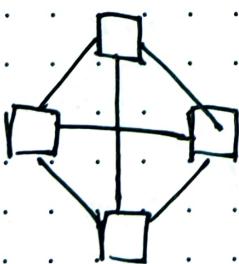
OO Integration Testing

- Related groups of ^{related} classes.

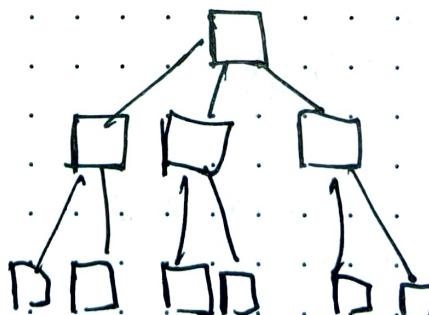
- Since architectural style changes, top-down and bottom-up approaches are irrelevant.

- stubs and drivers still exist though.

OO Architecture



Call & Return



Integration testing approaches:

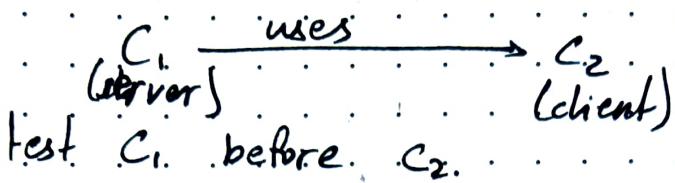
Thread-based

Classes handling similar events are tested together.

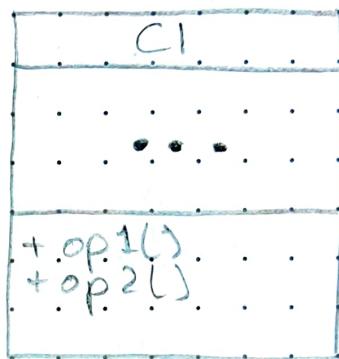
e.g. Student Registration, Course, etc. all handle registration events.

User-based approach

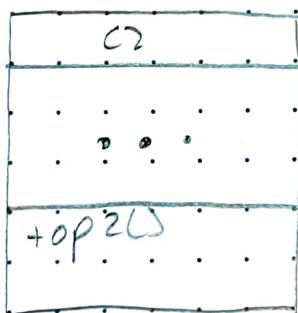
Start by testing the independent classes, and then proceed outwards.



Use test cases for C₁'s op1/op2
when testing C₂, and
write new test cases
for op3.

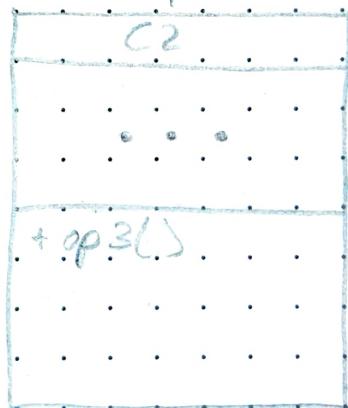


if C2 is



write new test cases
for op2 as well

if C2 is as above AND
C1's opt is



11. In this case.

we need to rewrite new test cases for op1 as well as op2.

Example of Unit Testing:

Unit: "BankAccount" class

minimum possible sequence:

open → deposit → withdraw → close

Anything can come

here, e.g. as below,

→ Random Testing:

open → deposit { deposit → deposit → balance → statement →
withdraw } withdraw → close

add any random sequence of valid function calls
between the boundaries.

→ Partition-based testing:

Partitioning approaches:

↳ Categories:

group operations into conceptually categories
e.g., initialization, termination, computation, query

↳ State-based:

partition based on whether the state has
changed or not

↳ Attribute-based:

for each attribute, create 3 partitions:
- functions that change its value
- functions that simply use its value
- functions which do neither.

SQE 13-11-2024

Software Configuration Management (SCM)

It is an umbrella activity: it spans the entire SDLC.

Software Configuration = Set of Software Configuration Items
=> SC = Set of SCIs.

SCIs

documents (SRS, test plan, test-case plan, etc.)
code

SCIs evolve over time, hence SCIs have different 'versions', hence the Software Configuration also have 'versions'.

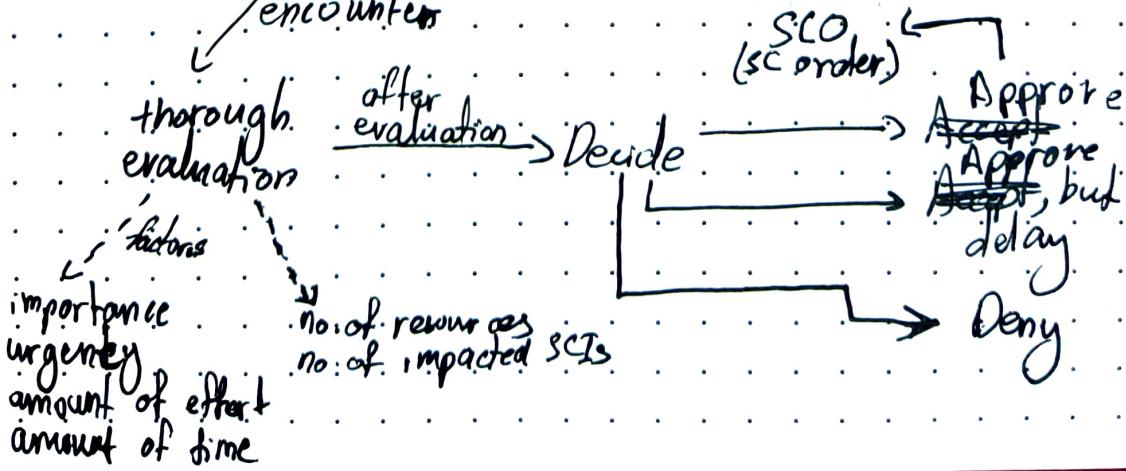
SCM Tasks

Change Management

We should not 'let' change happen, changes must be controlled to avoid derailing the project. Further, changes must be accompanied by regression testing.

Software Change Request (a.k.a Change Request)

upon encounter not an 'order'



Depending on contractual clauses, the customer may be charged for changes.

Change Management is overseen by a department called Change Control Authority/Board (CCA/ccb).

Release Management

A software configuration (sc) is released during deployment/delivery.

Types of versions

- Baseline: correspond to milestones, and are pre-planned.
- Intermediate: added as a result of one or more change requests

Baseline version
1

Baseline version
2



Software Configuration Management Plan:

- how many baseline versions?
- when?
- contents / impacted SCIs in baseline versions?

<Software> Ver. <Version #> <Revision #>

↳ technically called
"decimal enumeration"

↳ decimal point

)
minor changes usually
done in intermediate
versions

SCI	Flex Ver. 1.0	Flex Ver. 1.1	Flex Ver. 2.0	F. Ver 2.1
SRS	Ver 1.3	Ver 1.4	Ver 1.5	Ver 1.5
Class D.	-	-	Ver 1.1	Ver 1.2
State D.	-	-	Ver 1.2	Ver 1.3
M1 code	-	-	-	-
M2 code	-	-	-	-
M3 code	-	-	-	-
Test Plan	-	-	-	-
User Manual	-	-	-	-

It is possible for SCI versions to remain unchanged across different configurations.

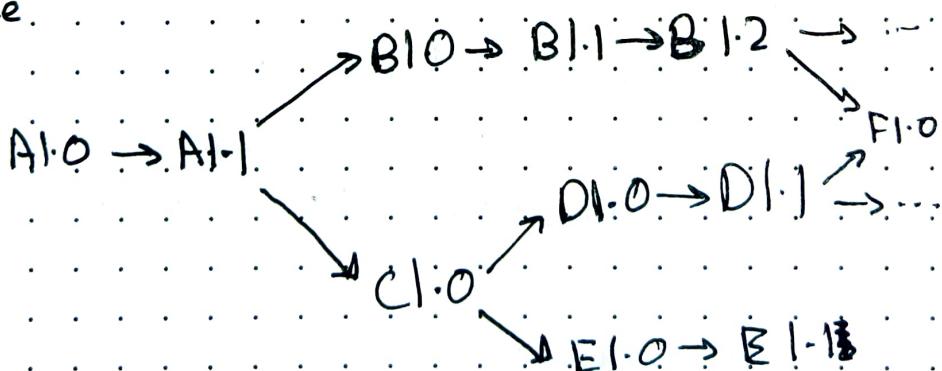
Software Evolution Approaches

- Linear

A1.0 → A1.1 → A1.2 → A1.3 → A2.0 → A2.1 → ...

- no branches
- used for single client or when only the latest versions & needs to be supported.

- Tree



- branches are multiple versions
- branches may merge, but continuing development will depend on the scenario.
- all leaf nodes are supported

- Provision of Information (SCM Task)

Version	Description	Release Document
-	- changes	
-	- open issues/known issues	
-	- deployment info (location, and version)	
-	- etc	

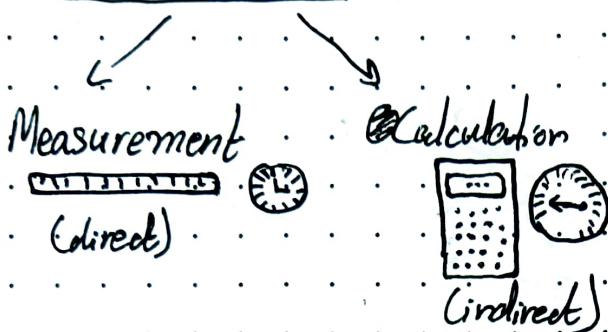
Auditing

- verify compliance to SCM procedures.
 - e.g. how many unapproved SCRs were implemented?
 - e.g. how many SCOs were delayed

Measurement Theory

Measurement: A process of assigning numbers/symbols to attributes of entities using certain rules.

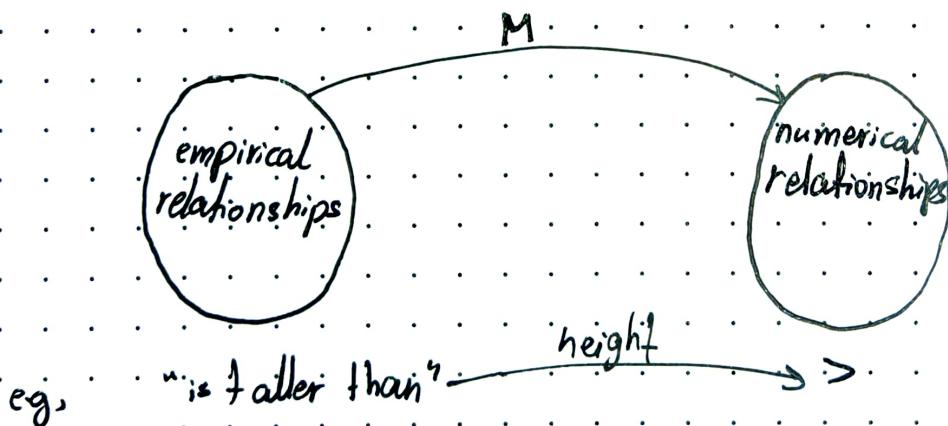
Quantification:



If you want to understand, control, or improve something you must measure.

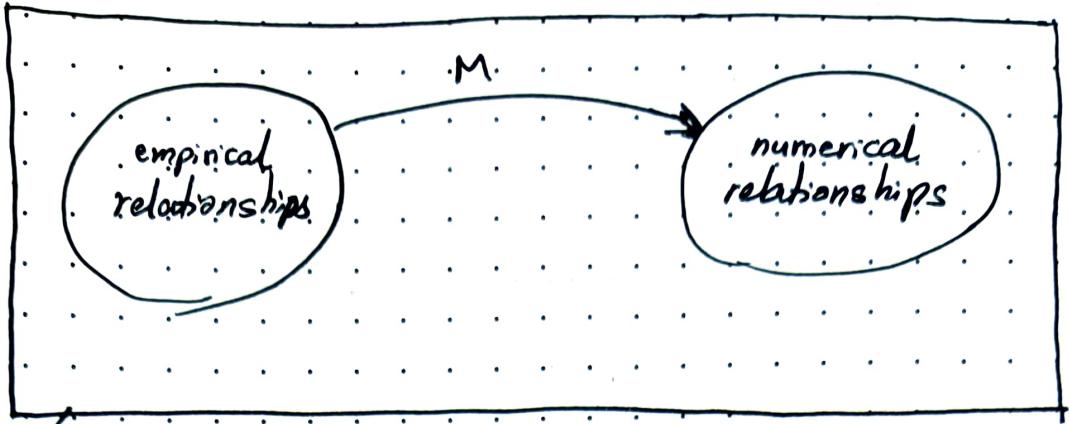
As if it wasn't formal enough, here's a more formal definition:

Measurement is a mapping between the real world and the formal mathematical world.



A is taller than B iff $M(A) > M(B)$

representation condition: a measure is valid if it satisfies the representation condition



- context; the measurement scale.
- measurement produces measurement(s).

Measurement Scales

- Nominal scale:

classes where no class precedes another
e.g. lions, camels, monkeys → animals

a b c d e f g h i j n

abc → lions
de → camels
fgh, ij → monkeys

- Ordinal scale:

classes but ordered.

e.g. all 'M' T-shirts are smaller than all 'L' T-shirts
The difference, interval, Δ (delta) need not be constant.

- Interval Scale:

- the difference/interval(delta (Δ)) between any two classes is constant.
- e.g., temperature in °C or °F
- zero quantity is not absolute zero.

- Ratio Scale

- the ratio between any two classes is constant.
- zero quantity is actual zero (absolute zero)

- Absolute Scale

- no. of % (anything)
- counting

Classification of Measurement Scale

Nominal	}	Qualitative
Ordinal		
Interval	}	Quantitative
Ratio		
Absolute	}	

Arithmetic Operations

- Nominal, Ordinal \rightarrow No arithmetic operations
- Interval \rightarrow Addition, Subtraction
- Ratio, Absolute \rightarrow All arithmetic operations.

Measurement of Central Tendency

- Nominal \rightarrow mode
- Ordinal \rightarrow mode, median
- Interval, Ratio, Absolute \rightarrow mode, median, mean

Admissible Transformation

Nominal	one-to-one
Ordinal	monotonically increasing transformation
Interval	$M' = aM + b$ (affined transformation) ($a > 0$)
Ratio	$M' = aM$ (ratio transformation) ($a > 0$)
Absolute	$M' = M$ (identity transformation)

Nominal:	Ordinal	Interval	Ratio	Absolute
$M \rightarrow M'$	$1^{\text{st}} \rightarrow 1$	$F = \left(\frac{9}{5}\right)C + 32$	Inches $= 12 \times \text{Feet}$	only 1 way to measure
Monkey $\rightarrow X$	$2^{\text{nd}} \rightarrow 4$			
Morse $\rightarrow Y$	$3^{\text{rd}} \rightarrow 200$			
Camel $\rightarrow Z$	$4^{\text{th}} \rightarrow 40000$			

SQE : 20-11-2024

Software Quality Measures and Metrics

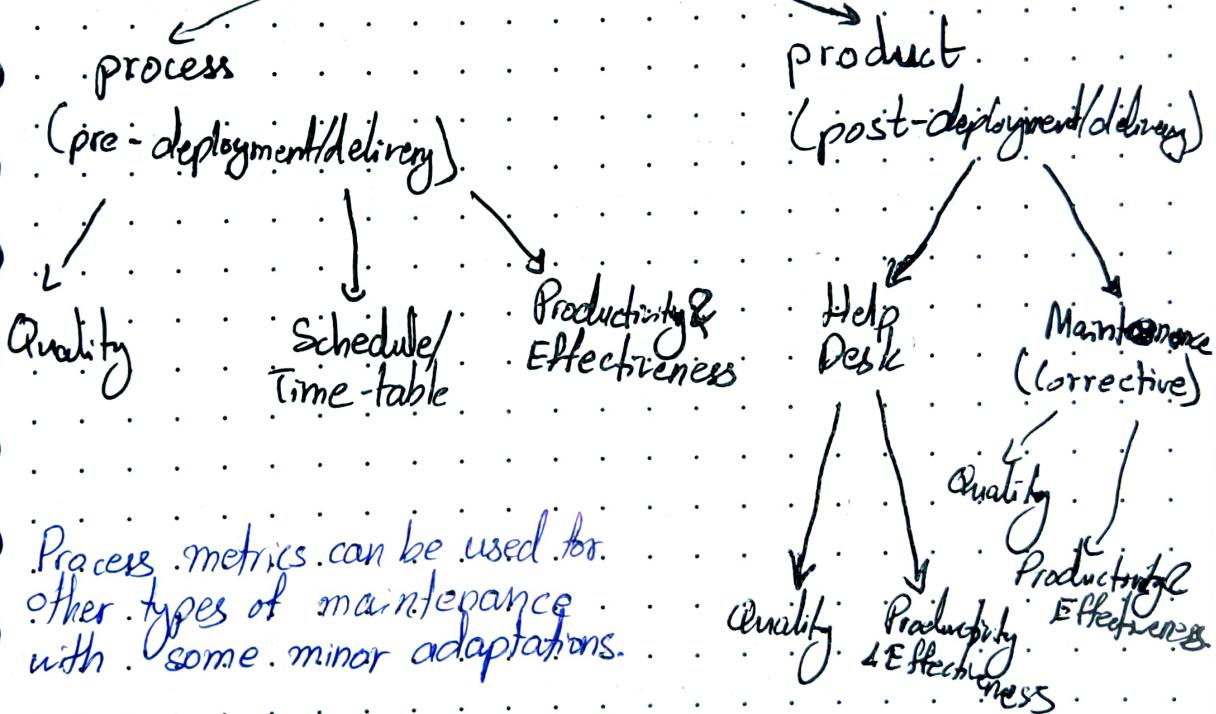
measured

calculated

Objectives

- Understand: understand what's going on
- { direct - Control: if things don't go as planned, controlling ~~steps~~ steps are taken.
- indirects - Improving: implement improvements

Measures & Metrics



Attributes for Successful Metrics

- Relevant: should measure attributes of significance
- Valid: should actually measure what they are designed to measure (satisfy the representation)
- Mutually Exclusive: ~~not to~~ measure something that has not been measured by another metric
- Reliable: produce similar results under similar circumstances
- Comprehensive: should not have limited applicability

- Easy & simple: easy to measure/calculate
- No independent data collection
- Immune to biased intervention

Size Metrics are used to normalize data. Comparisons can only be made once data is normalized:

Size Metrics:

- KLOC: kilo lines of code
= lines of code $\frac{1000}{1000}$
 - code counters can count lines automatically.
 - rules: comments/blank lines
 - objective value
 - value depends on the programming language
 - also somewhat dependent on programmer style: not immune to biased intervention
- FP: function points:
 - require human intervention, subjective
 - functionality remains the same irrespective of the programming language
 - not influenced by programmer style
 - user-stories/story points
 - class-points

A	B
100 med. faults	1000 med. faults

Quality of A and B can be compared

int	5 physical LOC
i = 0 ;	1 logical LOC

IFPUG: International Function-Point User Group

- function points can be calculated from the detailed SRS, hence they can be used for project estimation.

Calculating Function Points Values (Overview)

1- CFP (Crude Function Points)

- a.k.a. Unadjusted Function Points

Components:

- User Inputs | UI
- User Outputs | UO
- User Online Queries | UCQ
- Logical Files (DB tables) | LF
- External Interfaces | EI

- Each component is weighed on an ordinal scale:

simple medium complex

2- RCAF (Relative Complexity Adjustment Factor)

$$RCAF = \sum_{i=1}^5 S_i \quad \text{where } S_i = 0, 1, 2, 3, 4, 5$$

$$0 \leq RCAF \leq 1.0$$

$$3- FP = CFP(0.65 + 0.01RCAF)$$

Gearing /
Conversion
Factors/Ratios
QSM

For attend - master in Java:

$$FP = 85.86 \quad \text{Java Gearing Factor} = 53$$

$$FP = 85.86 \times 53 = 4,550.58 \text{ LOC}$$

$$\frac{4550.58}{1000} \approx 4.55 \text{ KLOC}$$

QSM CF language

$$CF_{language} = \frac{LOC}{FP}$$

$$FP \times CF_{language} = FP \times \frac{LOC}{FP} = LOC$$

$$KLOC = \frac{LOC}{1000}$$

ERROR DENSITY METRICS → Process Quality

Density = $\frac{\text{Mass}}{\text{Volume}}$ for software; volume = size

NCE = no. of code errors

NDE = no. of development errors

$$CE \subset DE \\ \Rightarrow NCE \leq NDE$$

WCE = weighted NCE

WDE = weighted NDE

CED = Code Error Density = $\frac{NCE}{KLOC}$

DED = Development Error Density = $\frac{NDE}{KLOC}$

WCED = Weighted CED = $\frac{WCE}{KLOC}$

WDED = Weighted DED = $\frac{WDE}{KLOC}$

WCEF = Weighted CE per FP = $\frac{WCE}{NFP}$

WDEF = Weighted DE per FP = $\frac{WDE}{NFP}$

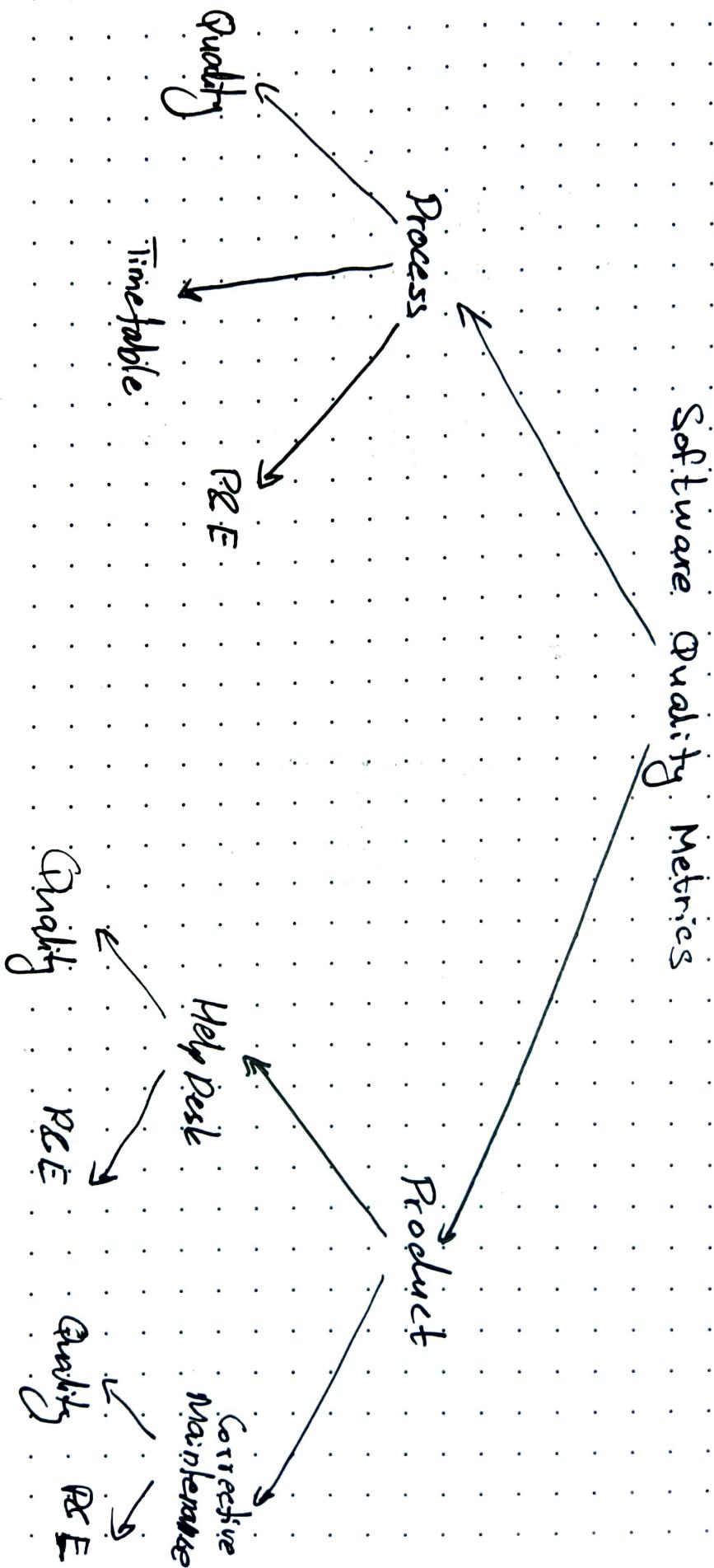
Example:

$$NCE = 70$$

simple	40	1	=	40
medium	20	3	=	60
complex	10	9	=	90
				190

→ WCE

QSQE 22-11-2024



Error Severity Metrics (Higher values bad)

$$ASCE = \frac{\text{Average Severity of Code Errors}}{\text{NCE}} = \frac{WCE}{NCE}$$

$$ASDE = \frac{\text{Average Severity of Development Errors}}{\text{NDE}} = \frac{WDE}{NDE}$$

Process Timetable Metrics

$$TTO = \frac{\text{Time Table Observance}}{\text{MS}} = \frac{MSOT}{MS}$$

$$ADM = \frac{\text{Average Delay of Milestone Completion}}{\text{MS}} = \frac{TCDAM}{MS}$$

MSOT = Milestones completed on time

MS = Total number of milestones

TCDAM = Total completion Delays (days, weeks, etc.) for all milestones.

ideally, $MSOT = MS$, and $TTO = 1$.
before-time completion is negative delay.

Error Removal Effectiveness Metrics

ideally, NYF should be zero.

Process Productivity Metrics

$$DevP = \frac{\text{Development Productivity}}{\text{KLOC}} = \frac{\text{Dev H}}{\text{KLOC}}$$

minimize DevP. (more Dev H/KLOC \Rightarrow less prod.)

$$FDerP = \frac{\text{Der H}}{\text{NFP}}$$

HD Calls Density Metrics (HD quality)

$$HDD = \text{HD calls density} = \frac{NHYC}{KLMC}$$

$KLMC = KLDC - \text{unmaintained code.}$

$NMFP = NFP - \text{reused part (unmaintained part)}$

Sensitivity of HD Calls Metrics.

$$ASHC = \frac{WHYC}{NHYC}$$

HD Success Metrics (quality)

$$HDS = \frac{NHYOT}{NHYC}$$

HD P&E

$$HDP = \frac{HDPYI}{KLMC} \quad (\cancel{\text{longer the better}})$$

HDE (lower the better).

Software System Failure Density Metrics

$$SSFD = \frac{NVE}{KLMC} \quad + \text{weighted} \\ + \text{FP variant.}$$

In Sensitivity Metrics

Failure of maintenance service metrics

$$MRepF = \frac{RepXF}{NYF} \quad (\text{lower the better})$$

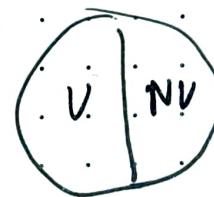
Software System Availability Metrics

FA \rightarrow full (vital + non-vital) Uptime / Down time

Vit A \rightarrow vital (only vital)

TUA \rightarrow none (no availability).

ideally, FA , $uFA \rightarrow 1$, $TCA \rightarrow 0$



$$NYFH \geq NYV_i + FH \geq NYTFH$$

$$\frac{N \times \text{SerH} - N \times \text{VitH}}{N \times \text{SerH}} \leq \frac{N \times \text{SerH} - N \times \text{VitH}}{N \times \text{SerH}} \leq \frac{N \times \text{SerH} - N \times \text{TFH}}{N \times \text{SerH}}$$

FA \leq VitA $\leq (1 - T(\text{VitA}))$

everything was
available.