Software Engineering Metrics, Experiments, and Models. Week 2.

Giancarlo Succi, Artem Kruglov

Innopolis University

January 12, 2023

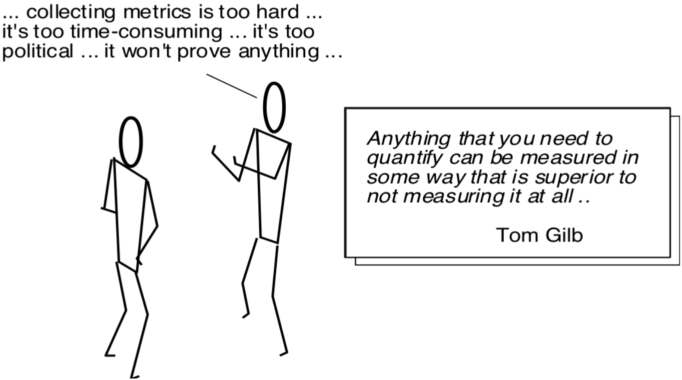
Measurement in Software Engineering

Content of Today and Next Week Lecture

- Introduction to Software Engineering Measurement
- Theory of measurement in software
  - Representational condition
  - Measurement scales
  - Measurements and models
- Examples of existing software measures
  - Size Oriented Metrics
  - Function Oriented Metrics
  - Quality Metrics
  - Other kinds of metrics

Introduction to Software Engineering Measurement

Measurement Metrics



from Pressman Companion Slides(chp 4)

Definition

Measurement is the process by which numbers or symbols are assigned to attributes of entities in the real world in such a way as to describe them according to clearly defined rules

N. Fenton and S. L. Pfleeger, 1997

Entities and Attributes

- Human being
- Apple
- Computer
- Computer
- Human being
- ...

- Name
- Weight
- Memory
- Hard Disk space
- Weight
- ...

Rules

- The LOC measure is obtained by counting all the lines that contain at least one character, and that do not contain comments.
- Interpretation: C Code?
- Visual BASIC 4 Code?

Software Engineering

- The collection of techniques that apply an engineering approach to the support and construction of software products

Fenton and Pfleeger, 1997

Why Measuring?

- Lack of measurable targets (Gilb's principle)
- Identification failure
- Lack of quality assurance
- Lack of consistent tool evaluation

The Measurement Advantage

Management

- Cost
- Productivity
- Quality
- User satisfaction
- Optimisation

Engineering

- Requirements testing
- Fault detection
- Meeting goals
- Forecasting

Scope of SW Measurement (partial list)

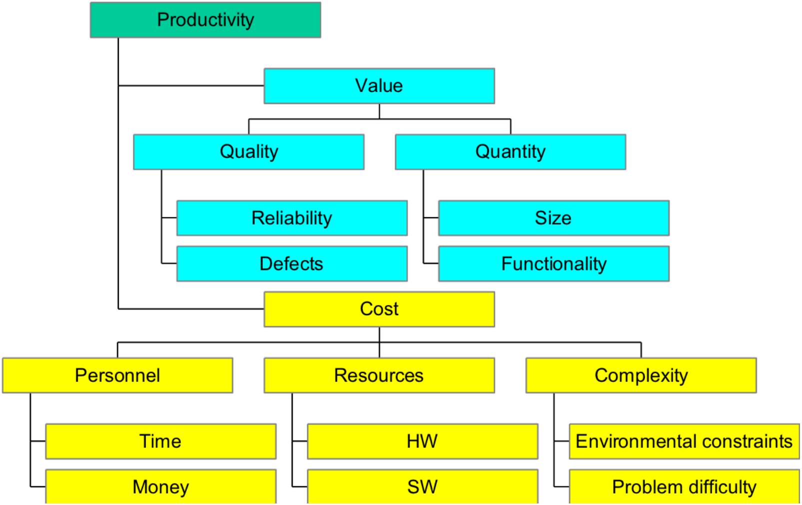
- Cost and effort estimation
- Productivity measures and models
- Data collection
- Quality models and measures
- Reliability models
- Performance evaluation and models
- Structural and complexity metrics
- Capability-maturity assessment
- Management by metrics
- Evaluation of methods and tools

Cost and Effort Estimation

- The aim is to predict project costs as early as possible
- Various models exist, expressing cost or effort as a function of size and/or other project attributes
- E.g., COCOMO (Boehm, 1981), COCOMO2 (Boehm, 1995), SLIM (Putnam, 1979), Function Points (Albrecht, 1979), ...

Productivity Models and Measures

Productivity Model (Fig. 1.3 of FP, pag. 15)

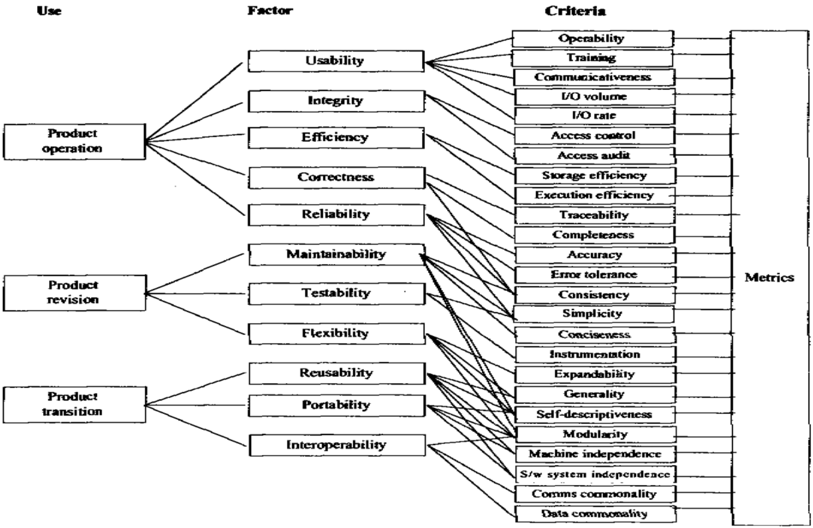


Data Collection

- Where
- How
- When

Should we collect software measurement data?

Quality Models and Measures



Reliability Models

- For how long will be our application up?
- Once we fix this failure, when the next failure will occur?

Performance Evaluation and Models

- Aspect of quality
- Include:
  - Response time
  - Average and worst case time complexity
  - Average and worst case space complexity
  - Required resources
  - ...

Structural and Complexity Metrics

- Describe inner properties of the system
- Examples include:
  - Weighted number of nested loops
  - Complexity of the program graph
  - Coupling between objects
  - ...

Management by Metrics

- Customers, developers, and managers can use software metrics to track the evolution of the project in terms of:
  - Hours spent
  - Required quality levels
  - Requests for new requirements
  - Work overtime
  - Quality of work

Evaluation of Methods and Tools

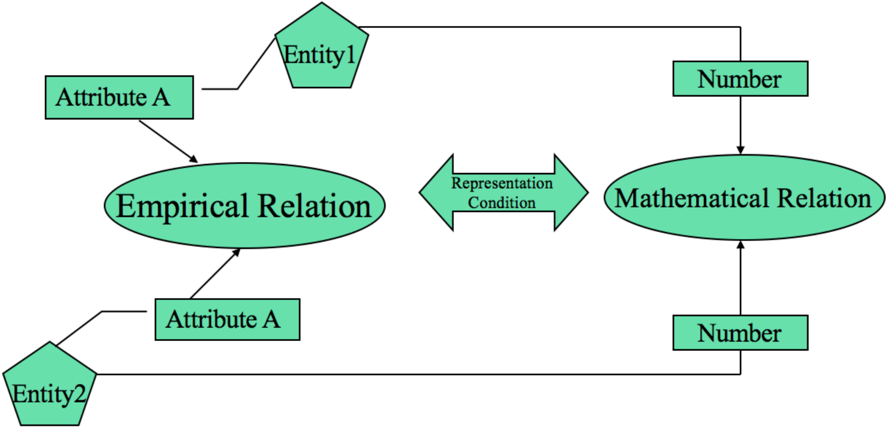
- New methods and tools can be very expensive for software companies both for:
  - Purchase / license
  - Training of employees
- It is therefore essential to have some systematic means to evaluate them (even if in most case this does not occur)
- Forms of experimental design help in defining proper structures for such studies

Capability Maturity Assessment

- Model of software production maturity
- Ensures customers of capability of the development process (e.g., ISO 9000 uses the buzzwords “say what you do and do what you say”)
- Provides paths for improvement
- Few exist ...

Theory of Measurement

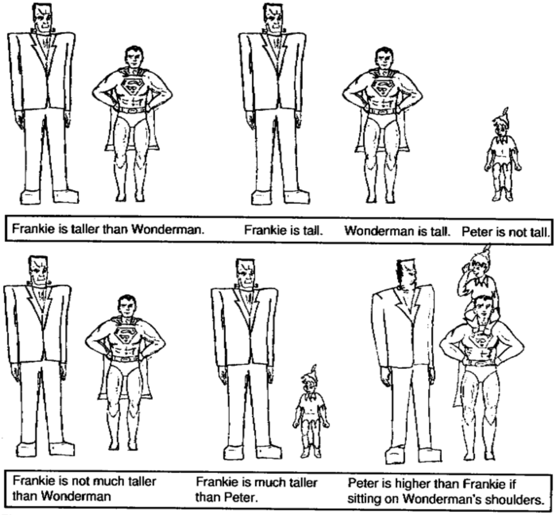
Set of rules to define measurement



Empirical Relations

- Links two entities by means of an attribute - e.g., “Person” (entity class) and “height” (attribute)
- Can be ambiguous - e.g., colour as perceived by human eye varies depending on subject measuring it
- Improves with the understanding of the attribute, and measures can foster improvement

Empirical Relations for the Attribute Height



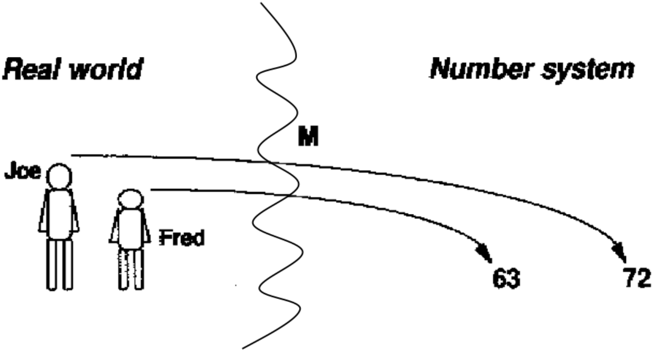
Measurement

To overcome these differences it is important to agree on:

- A **measurement** is a mapping from the empirical world to the formal, relational world
- A **measure** is the number or symbol assignment to an entity by this mapping in order to characterise an attribute

Fenton and Pfleeger, 1997

Measurement



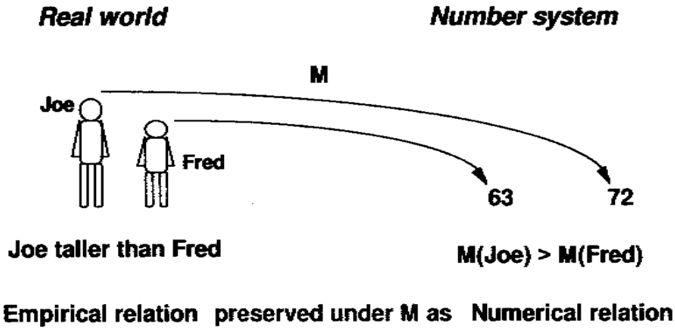
from Fenton pg 31

Representation Condition

- A measurement mapping must map entities into numbers and empirical relation into numerical relations that preserve them and vice-versa

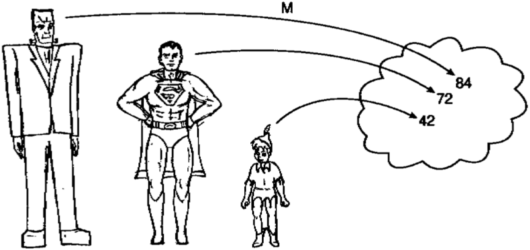
Fenton and Pfleeger, 1997

- **Valid** measure: satisfying the representation condition



from Fenton pg 31

Measurement Mapping



from Fenton pg 32

A tall iff $M(A) > 70$
A taller than B iff $M(A) > M(B)$
A much taller than B iff $M(A) > M(B) + 20$

Role of the Representation Condition

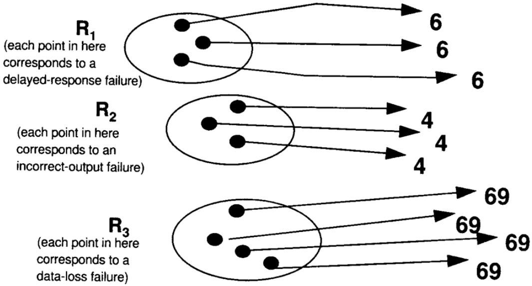


Figure 2.7: Measurement mapping

from Fenton pg 34

Alternative ...

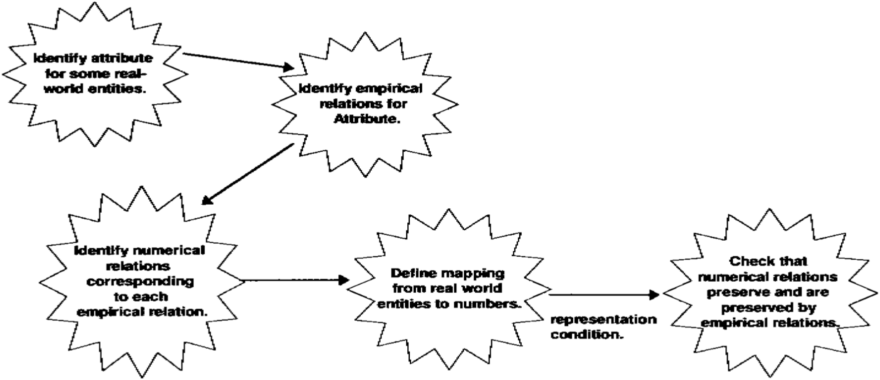
$M(R1) = 3$

$M(R2) = 4$

$M(R3) = 5$

Your Example ... :)

Key Stages of Formal Measurement



from Fenton pg 33

Mapping

- Attribute values – > numbers or symbols
- Attribute value domain – > range
- Empirical relation – > mathematical relation

Scales

Measurement Scales

- A measurement scale is a class of mapping that links empirical and number relations with specific properties

Measurement Scales

- Best possible numerical relation system?
- Representation of an empirical relation in a numerical system?
- Choosing a unique (and best) number system?

Measurement Scales

- Nominal (gender)
- Ordinal (arrival order)
- Interval (temperatures in F)
- Ratio (height)
- Absolute (the actual count)

Measurement Scales

- Language(Program) = 1, if Program is written in Pascal
- Language(Program) = 2, if Program is written in C
- Language(Program) = 3, if Program is written in Fortran

Few mathematical operations are applicable (mode, histograms, ...)

Measurement Scales

- Difficult(Program) = 1, if Program is easy to read
- Difficult(Program) = 2, if Program is not hard to read
- Difficult(Program) = 3, if Program is hard to read

We can have the median here...

Measurement Scales

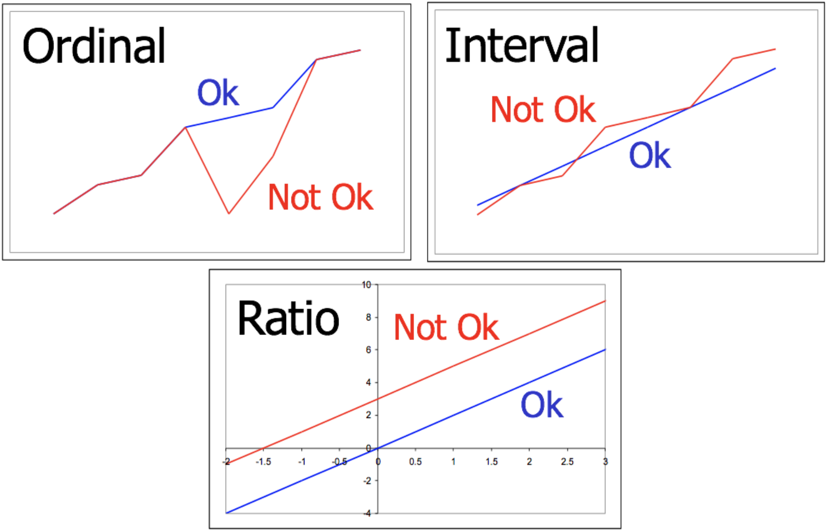
- Interval measures preserve differences but not ratios. Ex., The absolute time when an event occurred.
- Ratio measures preserve also the ratio between entities. Ex., LOC in a program. *All math operations are applicable.*
- Absolute measures are counts. Ex., the number of if statements in a program.

Table 2.8: Summary of measurement scales and statistics relevant to each (Siegel and Castellan, 1988

Scale type	Defining relations	Examples of appropriate statistics	Appropriate statistical tests
Nominal	Equivalence	Mode Frequency	Non-parametric
Ordinal	Equivalence Greater than	Median Percentile Spearman r Kendall τ Kendall W	Non-parametric
Interval	Equivalence Greater than Known ratio of any intervals	Mean Standard deviation Pearson product-moment correlation Multiple product-moment correlation	Non-parametric
Ratio	Equivalence Greater than Known ratio of any intervals Known ratio of any two scale values	Geometric mean Coefficient of variation	Non-parametric and parametric

Acceptable Mappings

- For nominal, any 1:1 mapping is OK
- For ordinal, the mapping needs to be strictly increasing
- For interval, the mapping must have the form $Y = aX + b$, with $a > 0$
- For ratio, the mapping must have the form $Y = aX$, with $a > 0$
- For absolute, the only acceptable mapping is $Y = X$


Ordinal
Ok
Not Ok
Interval
Not Ok
Ok
Ratio
Not Ok
Ok
10
8
6
4
2
0
-2
-4
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
3

Meaningful Measures

- Measures are said to be meaningful if their truth value does not change when the measure is subject to transformation
- That is, they are defined on the appropriate scale. Mapping is used to verify the appropriateness of the scale.

Examples

Meaningful

- The number of atoms in solid A is double the number of atoms in solid B
- The number of people who agreed was double the number of people who disagreed

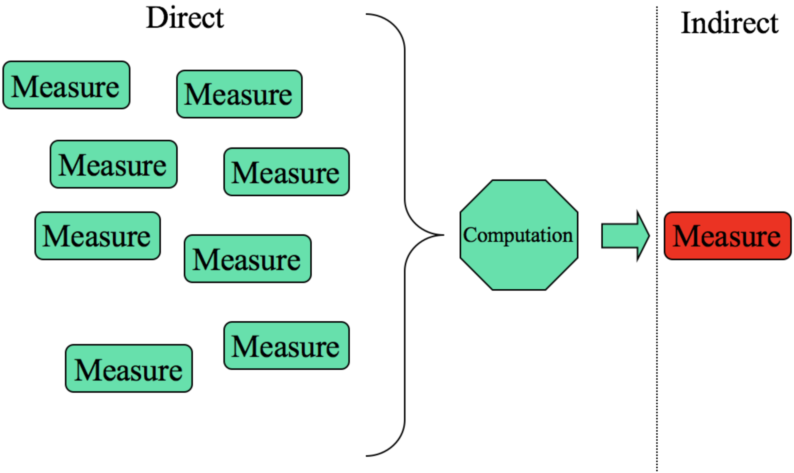
Not meaningful

- The color of solid A is twice as black as the color of solid B
- People agreed twice as much as they disagreed

Kind of metrics

- A metric is objective if it can be taken by an automated device; it is subjective otherwise
  - *LOC are objective metrics, Function Points are subjective*
- A metric is direct if it can be directly detected, indirect if it is the result of mathematical elaboration on other metrics
  - *LOC, number of errors, and FP are direct*
  - *Number of errors per LOC (Error density) is indirect*

Direct and Indirect Measurement



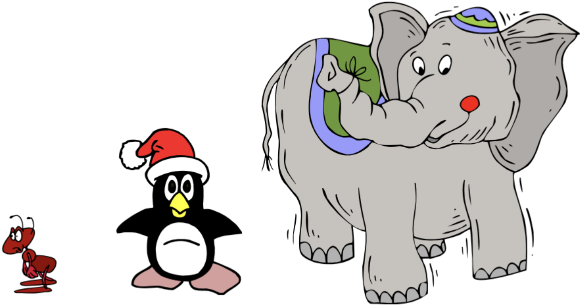
Direct or Indirect

- Immediately definable on one single calculation. Example: LOC, number of people in classroom, number of customer complaints

- Derived from a varied set of values. Example: ROI, number of tennis balls by weight, customer satisfaction

- Measurement scales limit the type of operations on measure - e.g., central tendency
- Objective or subjective measurement may limit the type of operations on measures
- Indirect measure depend on other measures' scales and thus are limited in meaningfulness and operations

Exercise: Measure of Mass

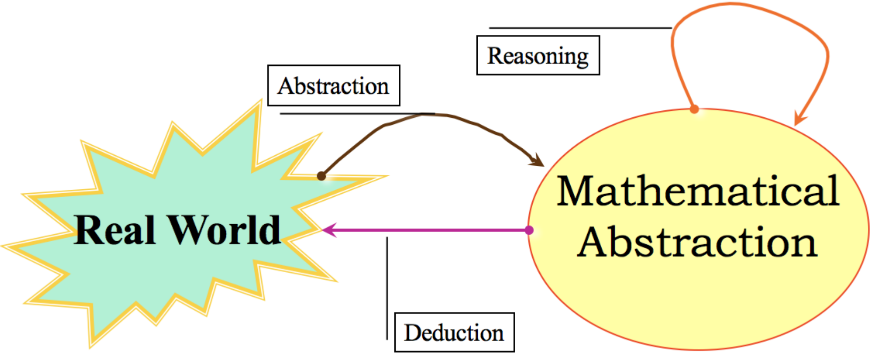


- What are he relations between their masses?
- Which of these are valid mappings?
  - $M_1(A) = 1,\ M_1(P) = 130,\ M_1(E) = 1400$
  - $M_2(A) = 3,\ M_2(P) = 4,\ M_2(E) = 5$
  - $M_3(A) = 24,\ M_3(P) = 51,\ M_3(E) = 49$
- Can we tell how intelligent they are from these mappings?

Questions

- Is it wrong to assert that “lines of code” is a bad software measure?
- What scale is used in “lines of code” measurement?
- Discuss the notion of “distance” in a vector space and its meaningfulness as a measure
- What kind of measure would you use for “program quality?”

- A baby should double its weight at the age of month 6.



Model

- Mathematical abstraction
  - Indirect measurement
  - Control measurement
  - Prediction measurement
- **Prediction system** couples a model with procedures that allow forecasting

Risks while building models

Figure 2.8: Using a suspect definition

from Fenton pp. 38