Measurement

Visualizzazione dell'Informazione Quantitativa

http://softeng.polito.it/courses/VIQ

SoftEng
http://softeng.polito.it

Licensing Note

Measurement

the process of empirical objective assignment of values to entities, in order to characterize a specific attribute thereof

Measurement

- Entity:
  - an object or event
- Attribute:
  - a feature or property of an entity
- Objective:
  - the measurement process must be based on well-defined rules whose results are repeatable

Terms

Measure (noun): variable to which a value is assigned as the result of measurement.

Measure (verb): Make a measurement.

Measurement: The process of assigning a number or category to an entity to describe an attribute of that entity.

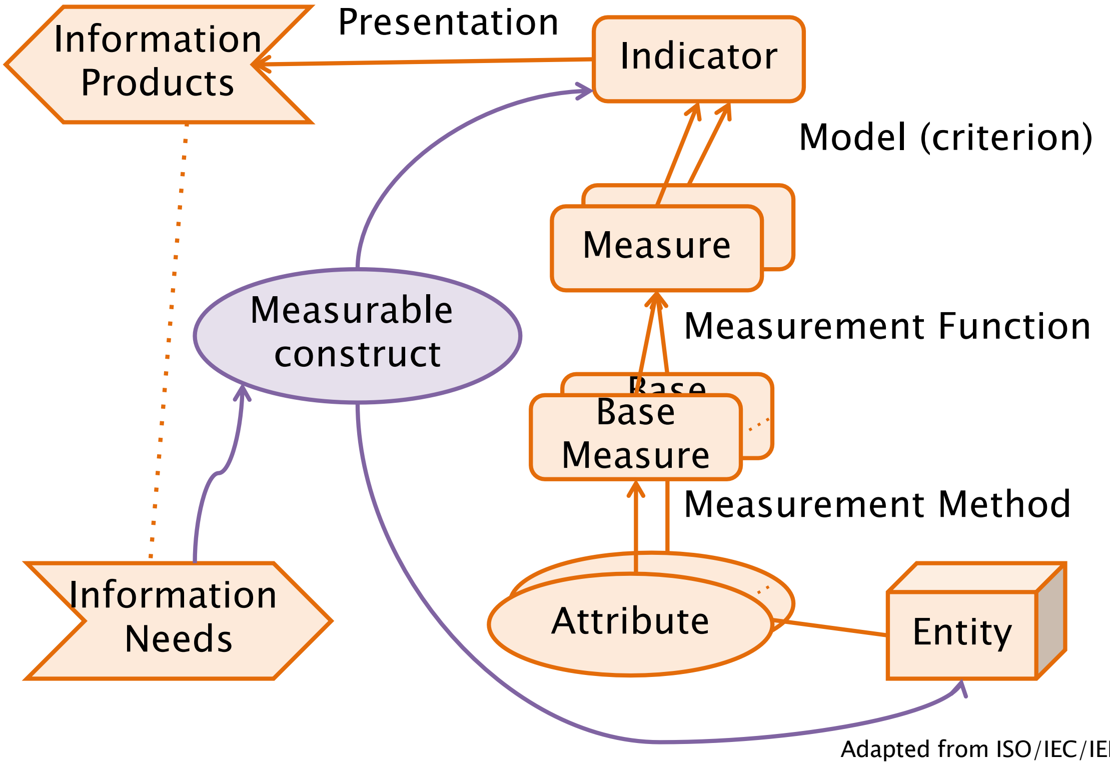
Metric: A measurement scale and the method used for measurement

Indicator: Measure that provides an estimate or evaluation derived from a model with respect to defined information needs

Construct vs. Measure

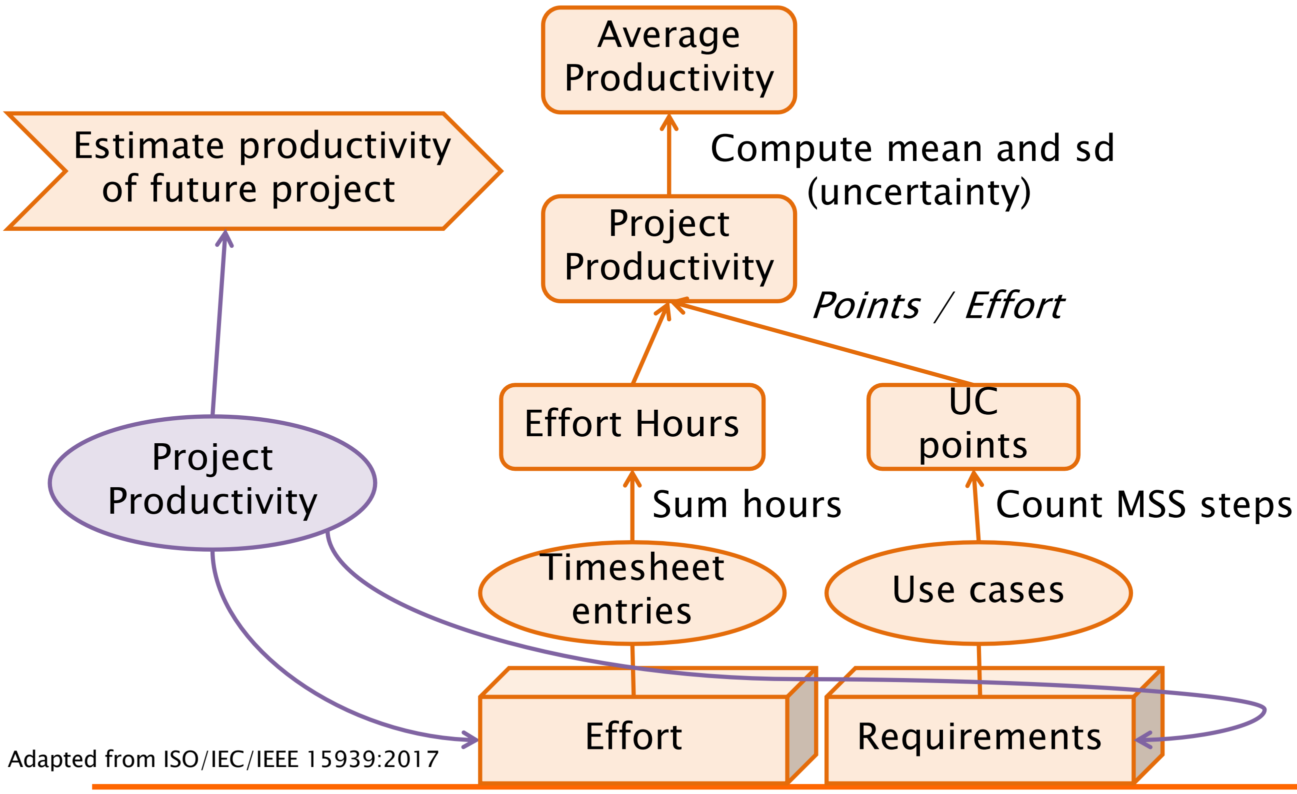
- Construct (conceptual)
  - Broad concept or topic of study
    - Abstract
    - Not (always) directly observable
    - Possibly ambiguous
    - May be complex (multiple aspects)
- Measure (Operational construct)
  - Precise definition
  - Clear measurement procedure

Measurement Conceptual Model



Adapted from ISO/IEC/IEEE 15939:2017

Measurement Example



Adapted from ISO/IEC/IEEE 15939:2017

Information needs

- The planning is driven by information needs dictated by other processes
- No (meaningful) measurement is possible without a clear goal

Examples of measures

Entity	Attribute	Measure
Person	Age	Year of birthday
Person	Age	Months since birth
Source code	Length	# Lines of Code (LOC)
Source code	Length	# Executable statements
Testing process	Duration	Time in hours from start to finish
Tester	Efficiency	Number of faults found per KLOC
Testing process	Fault frequency	Number of faults found per KLOC
Source code	Quality	Number of faults found per KLOC
Operating system	Reliability	Mean Time to Failure

Types of measures

- Base measure
  - Collected directly by interacting with the entity under measurement
- (Derived) Measure
  - Computed on the basis of other measures (either direct or indirect)

Base measures

- Length of source code
  - E.g. measured by LOC
- Duration of testing process
  - E.g. measured by elapsed time in hours
- Number of defects discovered during the testing process
  - E.g. measured by counting defects
- Effort of a programmer on a project
  - E.g. measured by person months worked

Derived measures

$$\text{Programmer productivity} = \frac{\text{LOC produced person}}{\text{months of effort}}$$

$$\text{Module defect density} = \frac{\text{number of defects module}}{\text{size}}$$

$$\text{Defect detection efficiency} = \frac{\text{number of defects detected}}{\text{total number of defects}}$$

$$\text{Requirements stability} = \frac{\text{\# of initial requirements}}{\text{total \#of requirements}}$$

$$\text{Test effectiveness ratio} = \frac{\text{number of items covered}}{\text{total number of items}}$$

$$\text{System spoilage} = \frac{\text{effort spent fixing faults}}{\text{total project effort}}$$

Internal vs. External

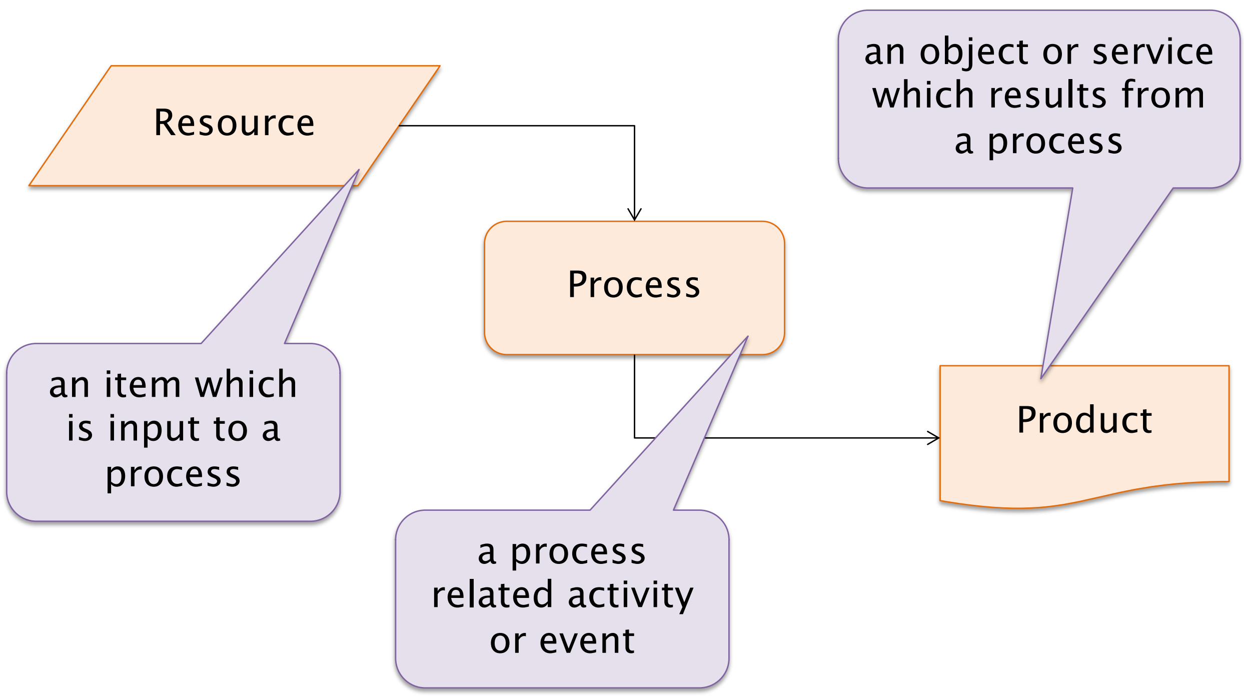
Given an entity:

- Internal measures can be measured purely in terms of the entity itself
  - e.g. length of source code (product)
- External measures can only be measured with respect to how the entity relates to its environment
  - e.g. reliability of application (product)
- Some attributes allow both internal and external measures
  - e.g. Usability
    - Internal: conformance to web standard
    - External: time to complete task

Objective vs. Subjective

- Objective measures do not depend on the environment or the person collecting the measure
  - Once you account for measurement error
- Subjective measures depend on the context where they are collected
  - Can change according to the person
  - They reflect the perception and judgment of the person performing the measurement

Entity classes



Sample measures

	Attributes	
Entities	Internal	External
PRODUCTS Specification, Source Code	Length, functionality modularity, structuredness, reuse	maintainability reliability
PROCESSES Design, Test	time, effort, #spec faults found time, effort, #failures observed	stability, cost-effectiveness
RESOURCES People, Tools	age, price, CMM level price, size	productivity usability, quality

MEASUREMENT THEORY

Evolution of measures

- More sophisticated measures can be defined as understanding of an attribute grows
- E.g. temperature of liquids:
  - 200BC: rankings, “hotter than”
  - 1600: first thermometer still “hotter than”
  - 1720: Fahrenheit scale
  - 1742: Centigrade scale
  - 1854: Absolute zero, Kelvin scale

Measurement theory

- Scientific basis to determine formally:
  - When we have defined an actual measure
  - Which statements involving measurement are meaningful
  - What the appropriate scale type is
  - What types of statistical operations can be applied to measurement data
- Based on foundation laid down by S.S. Stevens (1946)

Measurement theory

- Representation problem
  - The set of axioms under which measurement is possible
    - Prescriptive (rational judgment)
    - Descriptive (conditions for measurement)
- Uniqueness problem
  - How unique is the measure or scale

Measurement Scale

- Empirical system
  - Entities
  - Their relevant attributes
  - Relation (empirical)
- Mapping
- Formal system
  - Measures / values
  - Relation (formal)

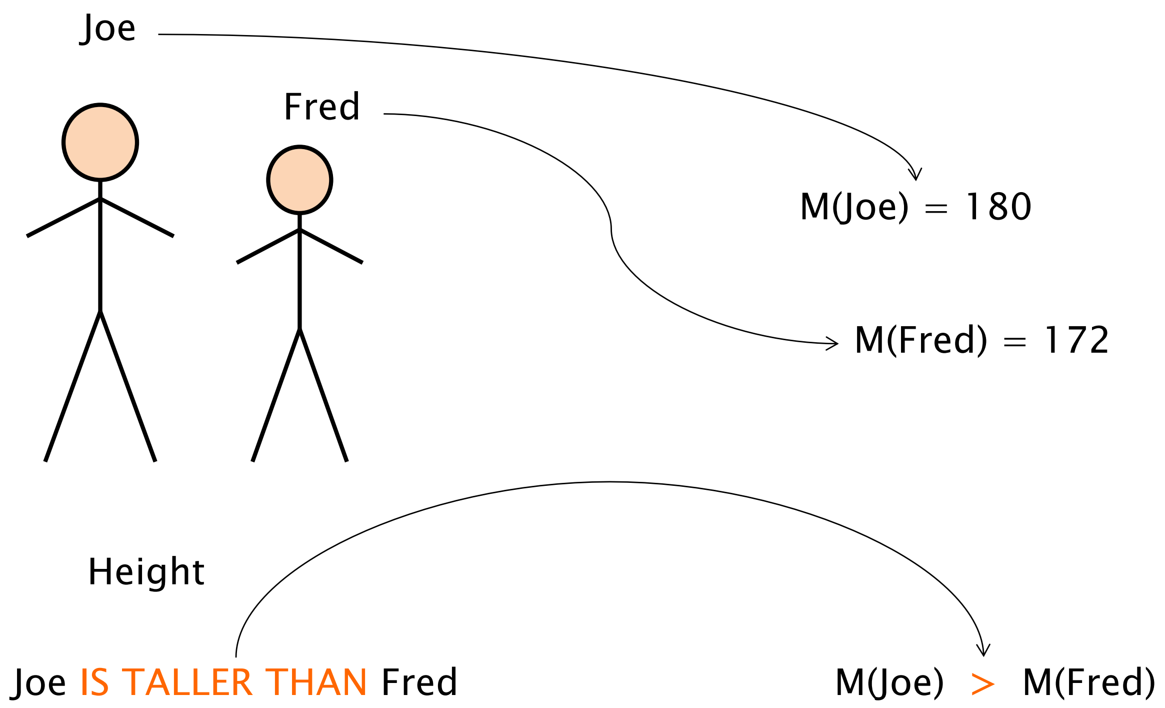
Empirical relation system

- A set of entities
- The relations that we observe among entities in the "real world" which characterize our understanding of the attribute in question
  - e.g. 'Fred taller than Joe' (for *height* of *people)*
- The closed operations that can be performed on the objects

Measurement mapping

- Mapping from the empirical relation system onto a formal relation system
- Consists of
  - Metric
  - Relation mapping
- A.ka. representation, homomorphism
- Measure: the value (formal element) assigned to an entity in order to characterize an attribute

Measurement mapping



Representation condition

- The measurement mapping implies that all empirical relations are preserved in formal (numerical) relations and no new relation is introduced
  - e.g. M(Fred) > M(Joe) precisely when Fred is taller than Joe
- Admissible metric if the representation condition holds
  - Measurement scale

Formally

- We can define a homomorphism *m*

$$
\begin{aligned}
\text{scale:} \quad & (\mathfrak{E}, \mathfrak{F}, m) \\
\text{empirical system:} \quad & \mathfrak{E} = (E, \text{taller}) \\
\text{formal system:} \quad & \mathfrak{F} = (\mathbb{R}, >) \\
\text{mapping function:} \quad & m : E \to \mathbb{R} | \\
& \quad \forall a, b \in E, a \text{ taller } b \implies m(a) > m(b)
\end{aligned}
$$

Additive metric

- A possible additional requirement is to have and additive measure

$$
\begin{aligned}
\text{scale:} \quad & (\mathfrak{E}, \mathfrak{F}, m) \\
\text{empirical system:} \quad & \mathfrak{E} = (E, \text{taller}, \text{added}) \\
\text{formal system:} \quad & \mathfrak{F} = (\mathbb{R}, >, +) \\
\text{mapping function:} \quad & m : E \to \mathbb{R} | \\
& \quad \forall a, b \in E : \\
& \qquad a \text{ taller } b \implies m(a) > m(b) \\
& \qquad m(a \text{ added } b) = m(a) + m(b)
\end{aligned}
$$

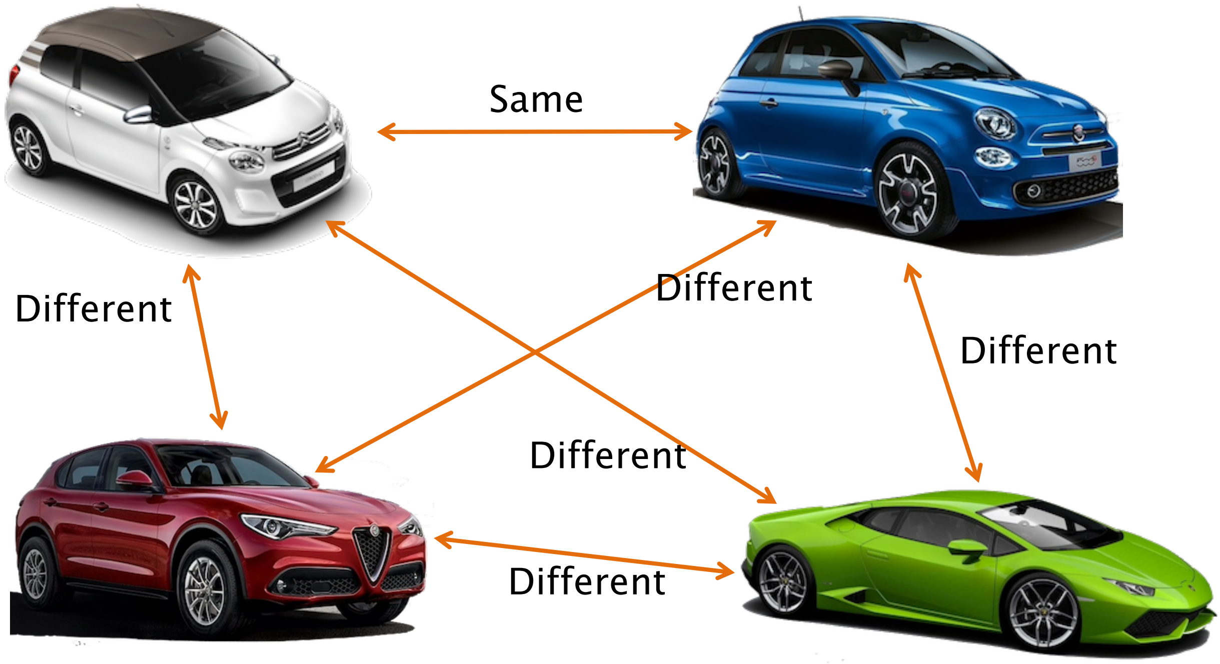
Admissible transformation

- Metrics are not unique, in general there are several homomorphisms
- Admissible transformation Φ
  - $\Phi \circ m$ is an homomorphism
  - Mapping between two measures, e.g. length
    - Admissible transformation: M' = a*M
    - Inadmissible transformation: M' = a*M + b

Relation System richness

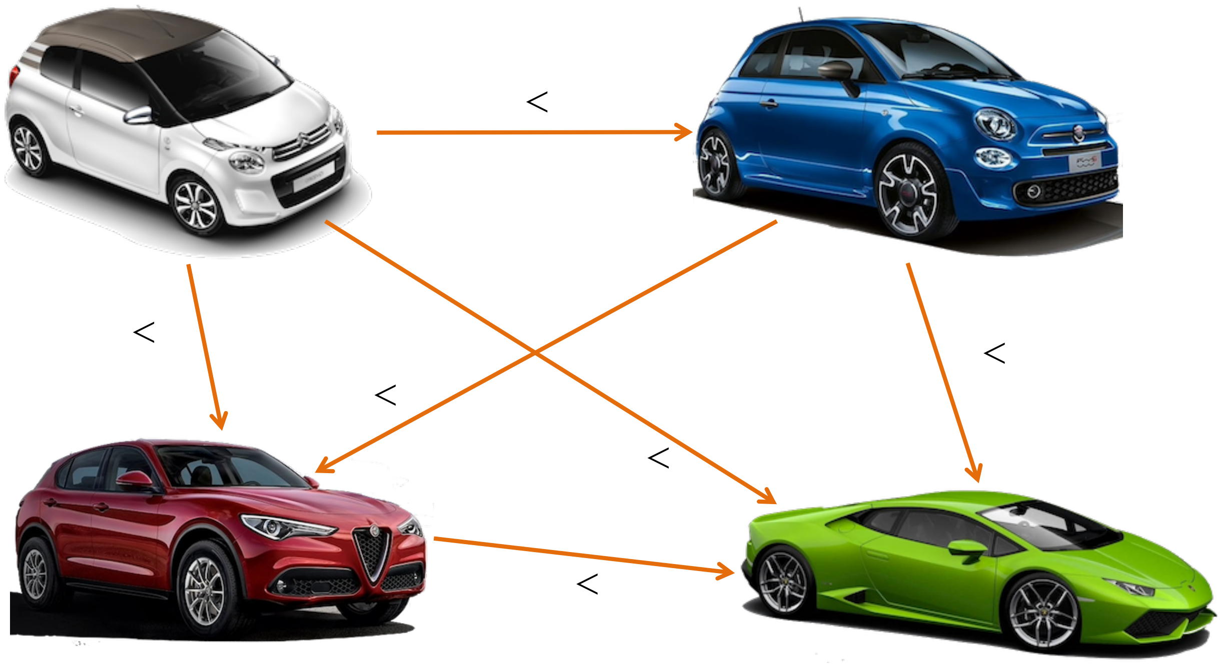
- RS_A is richer than RS_B if all relations in RS_B are contained in RS_A
- The richer the empirical system the more sophisticate the scale
- Complex and well understood phenomena require more sophisticate measurement scales

“Poor” Relation System



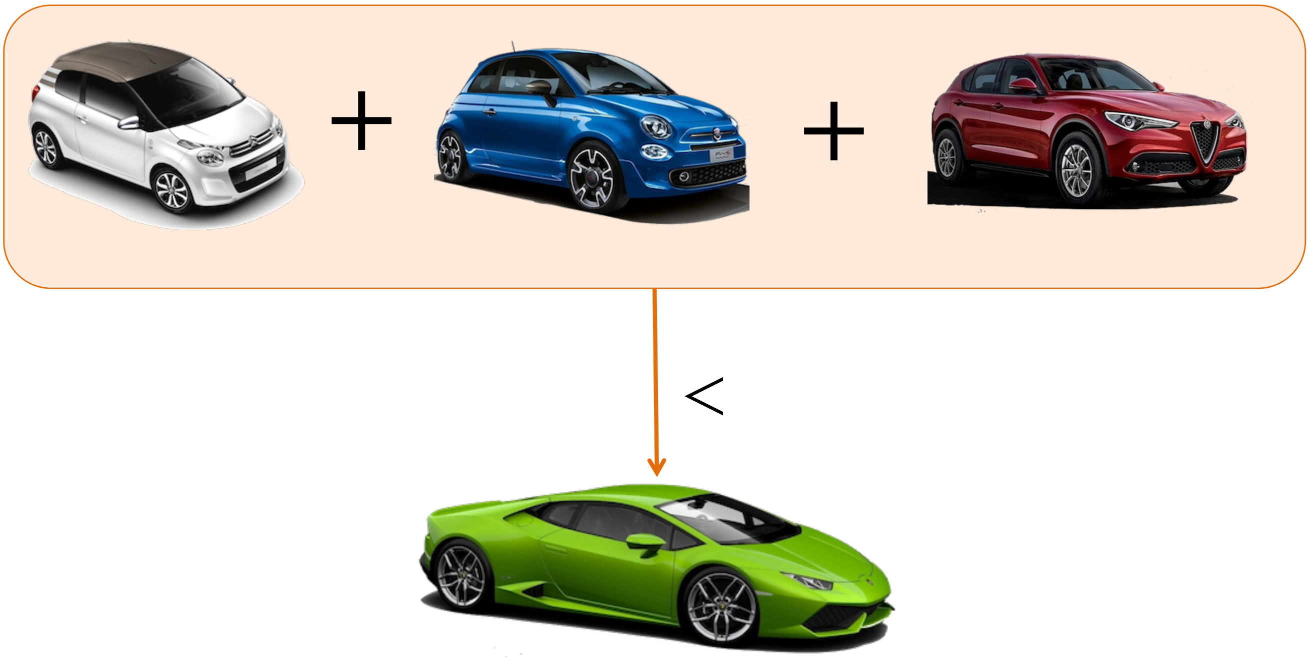
Attribute: category

“Lees poor” Relation System



Attribute: fuel consumption

“Rich” Relation System



Attribute: price

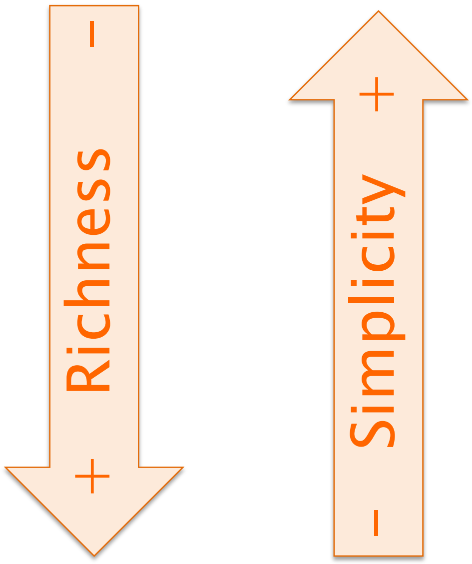
MEASUREMENT SCALES

Scale classification

- Measurement scales can be classified according to the class of admissible transformations
  - The larger the set of admissible transformations, the looser, less accurate, and less rich the scale
  - The smaller the set of admissible transformations the more accurate and richer the scale

Scale types

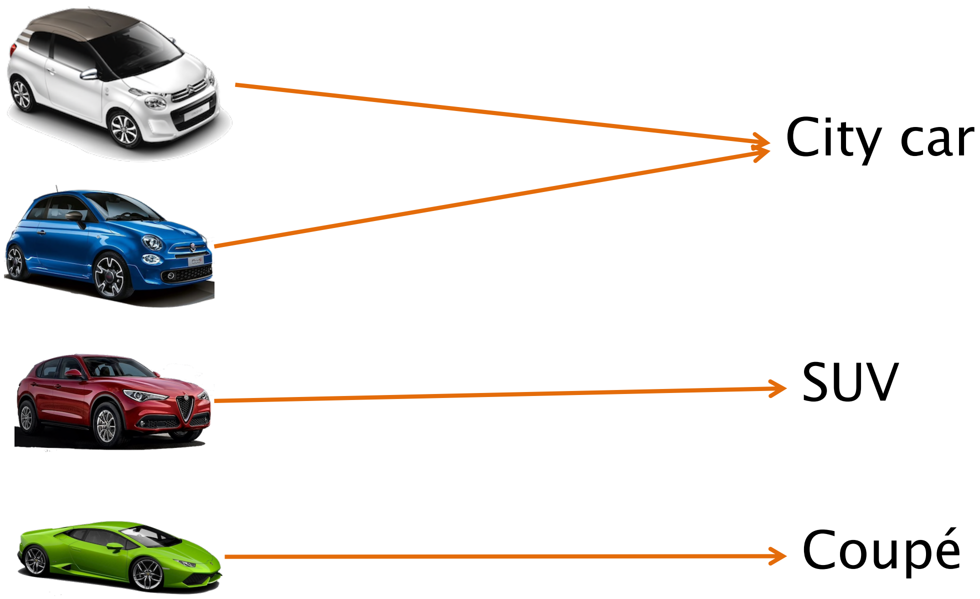
- Nominal
- Ordinal
- Interval
- Ratio
- Absolute



Nominal scale

- Places elements in a classification scheme
- Empirical relation: different classes
  - No ordering relation
- Any distinct numbering or symbolic representation is acceptable
  - No notion of magnitude

Nominal scale example



Attribute: category

Nominal scale example

- Empirical system
  - Entity: person
  - Attribute: origin
    - Italy, EU, Extra-EU
- Admissible mapping
  - M(p) =

I if *p* is from Italy
E if *p* is from any EU country
X if *p* is from a non EU country

Nominal scale example

- Empirical system
  - Entity: fault
  - Attribute: location
    - Specification, design, code
- Admissible mapping
  - M(x) =

S if x is a specification fault
D if x is a design fault
C if x is a code fault

Nominal Statistics

- Only a base operation: count
- Available statistics
  - Frequency (per category)
  - Mode

Ordinal scale

- Empirical system: classes of entities ordered w.r.t. attribute
- Empirical relation: total order
- Acceptable mapping: any mapping preserving the order
  - Measure represent ranking only
  - Acceptable transformations are the set of all monotonic mappings
  - $<C1, C2, .. Cn> \rightarrow <a_1, a_2, .. a_n>$
  - Where $\forall$ i>j, $a_i > a_j$

Ordinal scale example

Low

Low

Medium

Very High

Attribute: fuel consumption

Ordinal scale example

- Empirical system
  - Entity: statement
  - Attribute: agreement
    - Completely disagree, Mostly disagree, Mostly agree, Completely agree
- Admissible mapping
  - M(x) =

-2 if x is Completely disagree
-1 if x is Mostly disagree
1 if x is Mostly agree
10 if x is Completely agree

Ordinal Statistics

- Operations:
  - Counting
  - Sorting
- Available statistics
  - Frequency (per category)
  - Mode
  - Rank
  - Quantiles (Median)

Interval scale

- Empirical system: order and differences between classes
- Empirical relation: distance from a reference point
- Acceptable mappings: preserve order and difference
  - Addition and subtraction make sense
  - The ratio makes no sense
- Acceptable transformations are affine transformations
  - $M' = a * M + b$

Interval scale example

2015

2015

2017

2016

Attribute: year of introduction

Interval scale example

- Empirical system
  - Entity: activity
  - Attribute: start day
    - Gregorian calendar
    - Chinese lunar calendar
- Admissible transformation
  - EU start
    - Day 36 (Feb 5, 2019)
  - China start
    - Day 1 (First day first Month, the Year of the Pig)

  - $D_{EU} = D_C + 36$

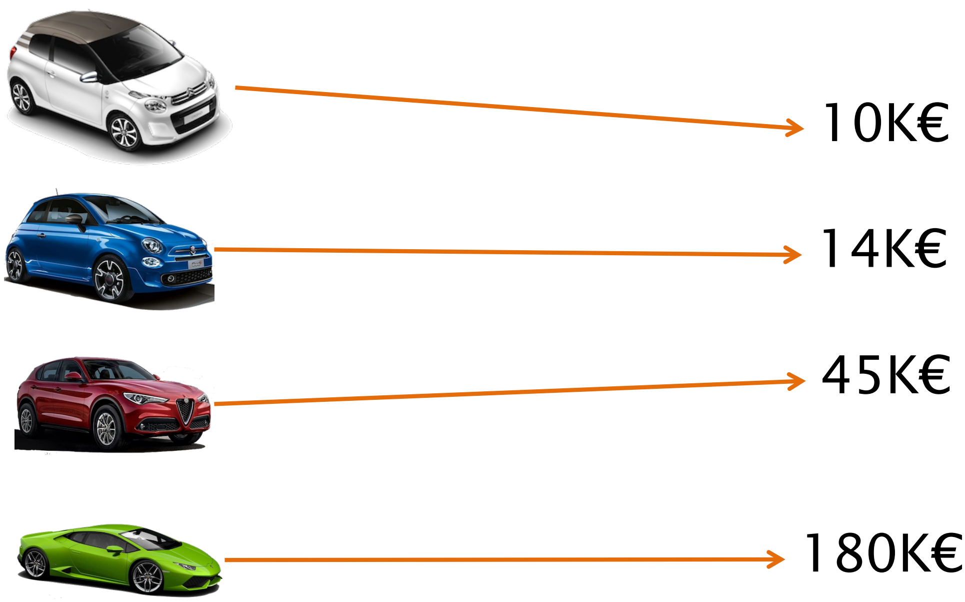
Interval Statistics

- Operations:
  - Counting, sorting
  - Sum, Difference, Scalar division
- Available statistics
  - Frequency, Mode, Rank, Quantiles
  - Mean (Arithmetic Average)
  - Variance (and derivatives)

Ratio scale

- Empirical system: there is a zero element
  - Represents total lack of attribute
  - Measurement starts at zero and increases at equal intervals (or part of): called units
  - All arithmetic can be applied meaningfully to classes in the range of the mapping
- Empirical relation: ratio between entities
- Admissible transformation
  - Ratio transformation
  - M' = a*M

Ratio scale example



Attribute: price

Ratio scale example

- Empirical system
  - Entity: person
  - Attribute: age
    - Years, Months
- Admissible transformation
  - $M_{Months} = a * M_{Year}$
    - Where a = 12

Ratio Statistics

- Operations:
  - Counting, sorting
  - Sum, Difference, Scalar division
  - Division, (Multiplication)
- Available statistics
  - Frequency, Mode, Rank, Quantiles, Mean (Arithmetic Average), Variance (and derivatives)
  - Standardized mean difference, etc.
  - Geometric mean, etc.

Absolute scale

- Measurement made by counting items in the entity set
  - Number of occurrences
  - Only one possible mapping
  - All arithmetic analysis is meaningful

Absolute scale (counter)examples

- Empirical system
  - Entity: project
  - Attribute: full time staff
    - Number of full time developers
- The attribute definition implies the items to be counted!
  - Length is not measurable on an absolute scale, # of lines it is
  - Age is not measurable on absolute scale

Scales

Scale	Admissible Transformations	Example
Nominal	1-to-1 mapping	Labeling, classifying entities
Ordinal	Monotonic increasing function	Preference, hardness
Interval	$M' = a*M+b$ With: $a>0$	Relative time, temperature
Ratio	$M' = a*M$ With: $a>0$	Time interval, length
Absolute	$M' = M$	Counting entities

Meaningful statements

- A statement involving measurement is meaningful if its truth is invariant of transformation to any admissible scale
  - i.e. the conclusion is the same after an admissible transformation is applied to the measures

Meaningful statements

- Statements
  - ✓ The number of errors discovered during the integration testing was at least 100
  - The cost of fixing each error is at least 100 ?
  - ✓ A semantic error takes twice as long to fix as a syntactic error
  - A semantic error is twice as complex as a syntactic error

Meaningful statements?

- Fred is twice as tall as Jane
- The temperature in Tokyo today is twice that in London
- The difference in temperature between Tokyo and London today is twice what it was yesterday

Statistical operations

- Central tendency

Type	Mean	Median	Mode
Nominal	✗	✗	✓
Ordinal	✗	✓	✓
Interval	✓	✓	✓
Ratio	✓	✓	✓
Absolute	✓	✓	✓

INDICATORS

Indicator

- Name
- Goal
- Metric definition
  - How computed
  - Unit of Measure
  - Scale
  - Aggregation
- Interpretation
- Source
  - where the data comes from

Interpretation

- *Only a Sith deals in absolutes*

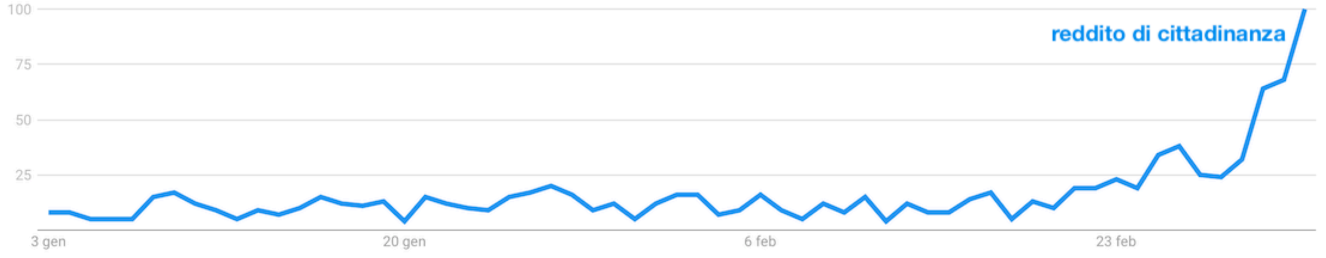
  – Obi-Wan Kenobi

- Interpretation entails a reference to:
  - Target
  - Benchmark
  - Time series
  - Population norm

Interpretation

- Conformance: compare to a specific business or usage requirement
- Benchmark: compare with a benchmark for similar product or system
- Time series: observe trend in time
- Population norms: compute quantile
  - Require a db of previous values

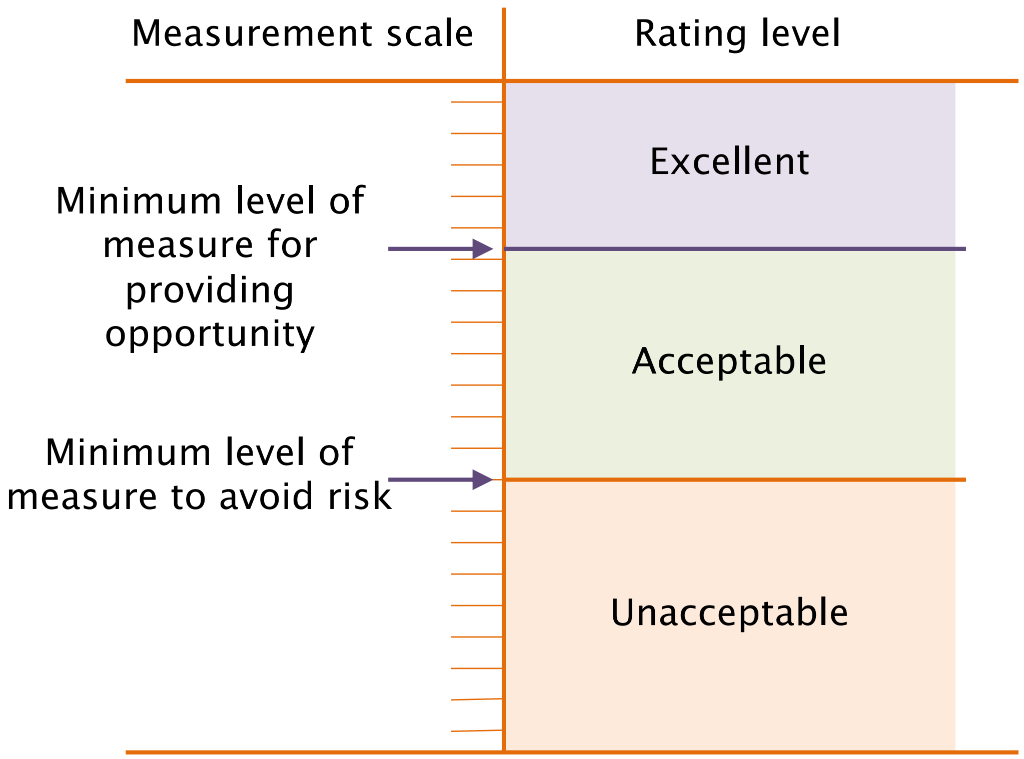
Interpretation requires context



https://trends.google.it/trends/explore?date=2018-01-03%202018-03-04&geo=IT&q=reddito%20di%20cittadinanza&hl=it&tz=-60

https://trends.google.it/trends/explore?date=2018-01-03%202018-03-04&geo=IT,IT&q=reddito%20di%20cittadinanza,di%20maio&hl=it&tz=-60

Interpretation: rating



Aggregation Dimensions

Or segmentation

- Entities to which indicator is associated and therefore
- Data the indicator can be aggregated on

- Dimensions are typically nominal or ordinal metrics

Common dimensions

- Time window
  - Sales per hours/per day/per month ..
- Hierarchical node in organizational – geographical structure
  - Sales per country/per region/per shop
  - Expenses per company/per division/per group/per person
- Product / product category
  - Sales per phone xy / per business phones
- Customer / customer category
- Activity in process
  - Cost per design / production
  - Defects from design/ from production
- Project
  - Cost per project
  - Defects per project

Robustness

- Understandability
- Processing cost
- Significance
- Frequency
- Structuredness

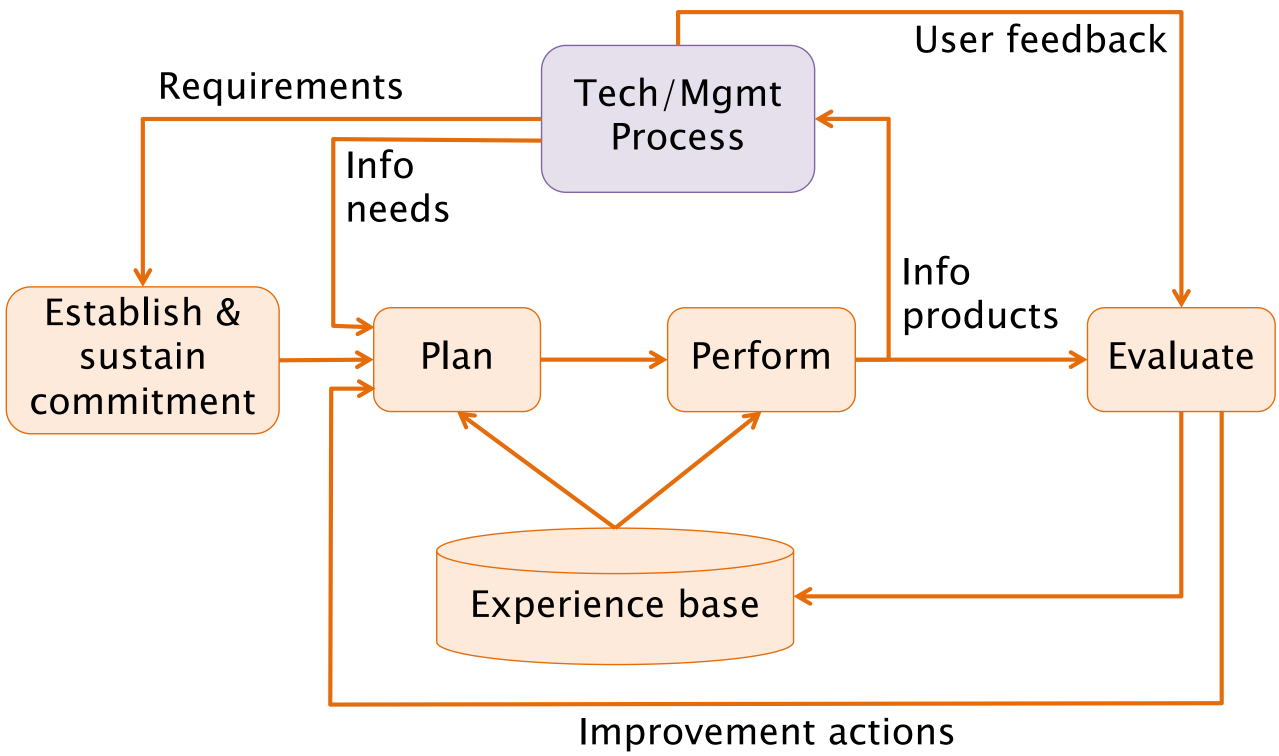
Indicator Robustness

- Comprehensibility / Understandability
  - How simple
- Processing Cost
  - Cost and delay to process
  - Cost and delay to collect raw data
  - Initial and incremental
- Significance / Meaningfulness
  - How much the indicator covers the information need
- Frequency
  - How often indicator varies
- Structuredness
  - How much the indicator is objective/not ambiguous

SMART Indicators

- Specific in purpose
- Measurable
- Achievable
- Relevant
- Timely

Measurement Process



Adapted from ISO/IEC 15939:2007

Measurement guiding principle

What gets measured gets done

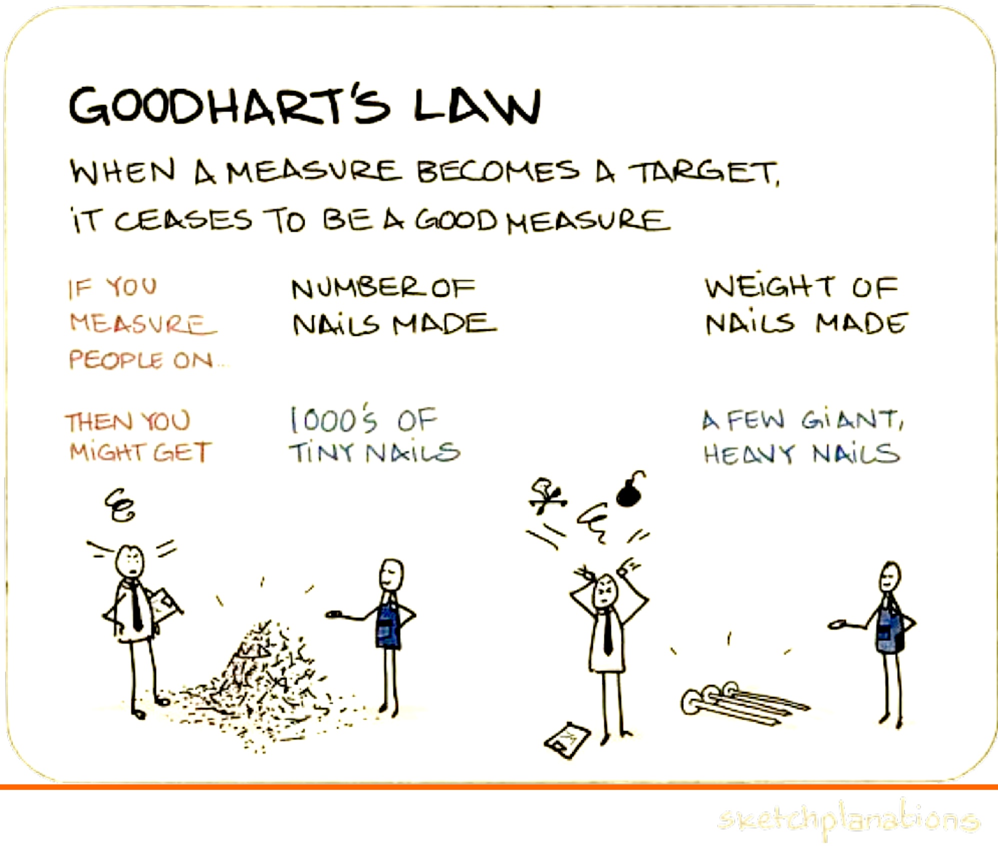
- Attributed to P. Drucker

Warning

The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor.

- Campbell's law

Goodhart's law



References

- Stevens, S. S. (June 7, 1946). "On the Theory of Scales of Measurement". Science 103 (2684): 677–680
- Roberts, F. (1979). "Measurement Theory with Applications to Decision Making, Utility, and the Social Sciences", Addison-Wesley.
- Fenton, N., Pfleeger. S. (1998) "Software Metrics: A Rigorous Approach". PWS Publishing.
- ISO (2007). Systems and software engineering — Measurement process, ISO/IEC 15939:2007(E)
  - http://ieeexplore.ieee.org/servlet/opac?punumber=4775908