MEASUREMENT AND ITS HISTORICAL CONTEXT

A thesis submitted
to Kent State University in
partial fulfillment of the requirements
for the degree of Master of Science

by

Naga Venkata Praveen babu Gedela

December 2008
Thesis written by

Naga Venkata Praveen babu Gedela

B.Tech., Cochin University of science and technology, 2003

M.S., Kent State University, 2008

Approved by

Austin Melton, Advisor

Robert A. Walker, Chair, Department of Computer Science

John Stalvey, Dean, College of Arts and Sciences
# TABLE OF CONTENTS

Acknowledgements ................................................. v

Dedication ......................................................... vi

Introduction ......................................................... 1

Mathematics ......................................................... 3
  Background ......................................................... 3
  Math modeling ..................................................... 5

Major Developments and People Involved in them ................. 8
  Euclid ........................................................... 8
  Hölder .......................................................... 8
  Campbell ......................................................... 11
  Stevens .......................................................... 15
  Suppes ........................................................ 18
  Luce and Turkey ............................................... 20

Fundamental Problems of Representational Measurement .......... 25
  Representation Theory ........................................... 25
  Fundamental Problems of Representational Measurement Existence .. 26

Software Measurement Meets Classical Measurement Theory .......... 27
Acknowledgements

My advisor Austin melton who guided me.
Your dedication here.

I dedicate it to my parents, brother and sister.
CHAPTER 1

Introduction

Measurement is important in every field of science. It is easy to measure attributes in some fields of science and difficult to measure them in some other fields of science. Software engineering is field of science where measurement theory is heavily used. Many software metric models were designed for different phases of software engineering like requirements, design, coding, implementing, testing and maintenance. Well known researchers like Norman Fenton focused measurement research in software engineering on the representational theory of measurement. [3] Representational measurement tries to understand what is being measured by collecting data from empirical observations and arranging them in some logical fashion in terms of familiar mathematical structures. Representational measurement is not just representation via numerals, but how entities relate to one another according to scientific theories like physics and physiology. The reason researchers in software measurement are interested in measurement theory is they believe that measurement theory principles, which have for the most part been developed for measurement in physical sciences, can help produce good software metrics. Interestingly, it should be noted that the measurement ideas of Norman Campbell who was sort of the “father” of modern measurement are not in complete agreement with the representation theory of measurement. Representational theory which seems to be the measurement theory most used in software measurement has to follow two basic requirements. The set of entities and there relations of a measurable attribute has to be similar to set of
numerals and there relations. They have to be isomorphic with each other. Difference is
relations on set of entities of an attribute have to be experimentally proved ex: ordinal,
additive. On other hand numerals are defined and may be conventional ex: 1, 2, 3...etc.
Unfortunately, users of software metrics often assume that every relation which exists on
set of numerals also exists on set of entities. Other basic requirement users ignore after
construction of measurement scale is that the measurement scale has to reflect the prop-
erties of set of entities and its relations. The applications which manipulate this scale
have to be in complete agreement with set of entities and its relation. But in software
measurement, software designers do not provide adequate attention to this requirement.
The same rules we apply in other fields of science may not work for software engineer-
ing. Software engineering is different enough from other fields of science like physics that
researchers need to develop a measurement theory which takes into account the needs,
requirements and conditions of software.

My thesis analyzes different measurement ideas from classical measurement theory and
representational theory and analyzes problems faced in software measurement by focusing
on examples from Halstead metrics.
CHAPTER 2

Mathematics

2.1 Background

Mathematics evolved from the need to do things like calculations, understand relationships between different static or dynamic objects and measure properties of objects. According to Steen mathematics is the study of patterns found in space, human beings, nature. Mathematics combined with logical reasoning and abstraction helped to translate empirical observations into systematic study. Today mathematics is used in different fields like economics, medicine, astronomy and physical science. Mathematics can be broadly divided into the four categories quantity, structure, space and change (arithmetic, algebra, geometry and analysis). In addition, we divide mathematics into applied mathematics and pure mathematics.

**Discrete mathematics:** Discrete mathematics is general term for the mathematics used in theoretical computer science. Discrete mathematics is the study of structures and is in some sense the opposite of continuous mathematics which includes computational complexity theory and information theory. Complexity theory studies the tractability of solutions for large complex problems. Information theory is the study of compression and entropy on large amounts of data on a given medium.

**Applied mathematics:** Statistics, probability theory and numerical analysis are important fields in applied mathematics. They provide tools to describe, analyze and predict problems in business, nature and space. The study helps in solving large mathematical
problems which are difficult to solve with just the human numerical capacity.

In any field, all four needs quantity, structure, space and change are interrelated. Quantity may lead to find the structure, or structure may lead to find quantity, or both may lead to analyze the change. We will discuss structure.

**Structure:** [15] Structure is a fundamental concept to describe or define anything like human body, nature, corporation, physical objects or psychology. It can be formally defined as set of entities and there relations. Relations can be many to many functions or one to many functions. Structures can be classified into mathematical structure (model) or abstract structure.

**Empirical Structures:** [3] We get data from empirical observations, experiments and measurements. Analyzing and arranging data systematically is called modeling data and can used in measurement theory. The data produce an empirical structure. An empirical structure from an experiment is a structured summary of the experimental activity. The data collected is regrouped into inputs, responses and relation between them. Usage of empirical structures is wide spread because they help organize data for subsequent statistical analysis or for testing a theory or hypothesis.

**Isomorphic structures:** [3] Isomorphism is mathematical concept used to identify identical structures of different objects or phenomena. If objects have same structure then they may have same properties. It helps to provide measurement methods or theorems unsolvable in each other structure. I will explain with an example how two structures can be isomorphic. Consider a structure with set of non empty elements or entities A. Let R be binary relations these elements have on each other. Let call it (A, R). Lets consider other structure with \( A' \) non empty elements and \( R' \) binary relation between
these elements. It is called \((A', R')\). These two structures are isomorphic with each other if and only if there is a function \(f\) such that 1. the domain of \(f\) is \(A\) and the co-domain of \(f\) is \(A'\), i.e., \(A'\) is the image of \(A\) under \(f\). 2. \(f\) is a one-one function. 3. For \(a\) and \(b\) in \(A\), \(a R b\) iff \(f(a) R' f(b)\).

**Ordered relational structures:** A structure has a set of elements or entities. They express relation with each other. One element can be greater or lesser than the other. This kind of structure is called weak ordered structure. It can be determined by two properties transitive and asymmetry.

Transitive: In set of elements \(A\) containing \(a, b, c\) if \(a\) is related to \(b\) and \(b\) is related to \(c\) then \(a\) is related to \(c\). Then it is said all these elements are transitive to each other. If relation is taken as \(\geq\) then \(a \geq b\) and \(b \geq c\) then \(a \geq c\).

Asymmetry: In set of elements containing \(a, b\) and have relations \(R\), for all \(a, b\) in \(A\) there \(a\) related to \(b\) but \(b\) is not related to \(a\).

Another question is, we have collected data from experiments, empirical observation and measurement tools. How should we use our mathematical concepts to extract information from this data? How to define a consistent methodology so that data collected is right and error free and information can be further utilized. The answer to this question is math modeling.

### 2.2 Math modeling

[4] We have quantitative data. We need understanding and insight as to how this data are connected. The answer to this question is a model. When data is quantitative, the
model is mathematical model which gives understanding and further information about the data.

We use measurement tools to extract information from data. Whenever you do an experiment, measure something or observe the universe you collect data. In this vast amount of data lies real information. Data can be anything, but information is precise. With scientific data, proof of understanding implies the capability to make accurate predictions. Qualitative conclusions are inadequate. Quantitative understanding starts with a set of well-defined metrics. Mathematical model are used in natural sciences, engineering, and social sciences. Models can be any form like differential equations, statistical models, dynamic systems, game theoretic models. A model can be defined with a set of equations and a set of variables. Variables can be anything like integers, real numbers, signals, events, time. Models can be classified into 1. Linear and non linear 2. deterministic and stochastic 3. Static and dynamic

**Data to Models**

Model is the medium between data and understanding.\[4\]

$\text{Data} = \text{Information} + \text{Error}. \ [4]$  

All data contains error. A good model identifies it and filters it. In its simplest form, a model is a filter designed to separate data into these two components. A mathematical model is a statistical filter.

Model can be constructed in two ways. The approach depends whether the information is known or the error is better known in data. If the properties of information are known we use that to extract information and leave error. This approach is commonly employed with stochastic data. Alternatively, if the error is the better known, the model
is designed to operate on the error, filtering it out and leaving the information behind. This approach is used on deterministic data. In both cases there will still be errors left. Designing a model with deterministic data is common place but a model with random variable data or stochastic data is difficult.
CHAPTER 3

Major Developments and People Involved in them

3.1 Euclid

Euclid Alexandria known as father of geometry was a Greek mathematician. The classical concepts of measurement were derived from Euclid V book. The major problem in those days was to identify measurable attributes. Euclid is first person to develop a theory to measure attributes. He introduced magnitudes. Magnitudes are whole numbers. Magnitudes are related with each other. A magnitude is a multiple of a smaller magnitude when it is measured by the smaller magnitude without remainder, i.e. the larger is composed of a whole number of copies of the smaller magnitude. A magnitude is a part of a larger magnitude when it measures that greater magnitude. Suppose M is a magnitude and \( k = 1, 2, 3 \) etc then \( kM = M \) if \( k = 1; M + (k - 1)M \), if \( k > 1 \). According to him, a measurable attribute has magnitude. All magnitudes sustain ratios. Attributes are measured in terms of ratio. For Euclid, ratio is sort of relation in respect of attribute between two magnitudes. For example, length is an attribute of a rod. If a rod is 2m and other rod is 4m. The second rod is 2 times more than first rod.

3.2 Hölder

Otto Ludwig Hölder was a famous German mathematician. He was famous for his classification of continuous quantity and discrete quantity with Hölder axioms and Hölder inequality. The issues faced during Hölder’s time were relations between magnitudes.
Euclid developed the theory of ratios of magnitudes without studying the nature of magnitudes. A magnitude is a part of another magnitude, the less of the greater, when it measures the greater. A ratio is a sort of relation with respect to size between two magnitudes of the same kind. For Euclid, relations were conceived as ratio of whole numbers [6]. But Hölder found there is a range of magnitudes for every attribute and they are related. He also found that magnitude is related with quantity. Let’s discuss quantity for a moment here. Quantity is a property of an attribute. It can be height, weight, long, short of an object. Magnitude exists on this quantity. Quantity can be continuous quantity or discrete quantity. Every quantity has range of magnitudes and there relation is shown through ratio. Examples of continuous quantity are mass, energy. Examples of discrete quantity are army, flock.

Quantitative Structure: Hölder is first to find out there is structure for continuous quantity. He described them using his seven axioms. The major features in his theory are range of magnitudes, relations, additivity, and continuity. Let explain them further.

Hölder Axioms:

Assumptions: a, b, c is magnitudes of an attribute, ‘+’ as operation and all magnitudes are positive.

1. a is equal to b (a=b) or not equal (a < b, a > b). It means the relationship between a and b has two states. They are equal or not equal.

2. For any lengths a and b \( a + b > a \). It means operation of addition between two magnitudes, the resulting magnitude obtained will be greater than individual magnitudes.
3. Order of operation doesn’t matter, $a + b = b + a$.

4. Additive relation is indifferent for compound operation $a + (b + c) = (a + b) + c$. It means the magnitude obtained by compound operation does not depend on order of operation. We get same result.

5. For any magnitudes $a$ and $b$ where $a > b$, there exist $c$ where $a + c = b$. It means for every two magnitudes, there exists few other magnitudes which form a relation with other magnitudes.

6. For every magnitude $a$ there exists magnitude $b$ less than $a$ ($b < a$). Where $b$ is less than $a$ but not equal to zero.

7. For each magnitude there exists a pair of non empty sets called a lower bound and upper bound. For each magnitude $a'$ with $a' < a$ then $a'$ is in the lower bound set. For each magnitude $a''$ with $a'' > a$ then $a''$ is in the upper bound set.

Conditions 2, 3, 4 define additive structure of magnitudes of an attribute. Hölder defines every magnitude of a quantity as measurable relative to any magnitude as a unit. For any magnitude “$a$” of attribute $Q$ there exists $2a = a + a$, $3a = (a + a) + a$, $4a = ((a + a) + a) + a$. $na = (n-1) a + a$. Given any pair of magnitudes $a$ and $b$ and $m$ and $n$

1. $m/n$ is lower fraction in relation to $a: b$ if $na > mb$.

2. $m/n$ is an upper fraction in relation $a: b$ if $na < mb$.

Ratio (measure) of magnitude $a$ relative to unit $b$ is determined via ordered sequence of (positive) real numbers, i.e. numerical ratios. Each positive real number has a least
upper bound non-empty set called a cut. It is also called a Dedekind cut. Each ratio of magnitude \(a:b\) contains a set of lower fractions and upper fractions and each contains its own well defined Dedekind cut. Hölder proved that a system of ratios of magnitudes of an unbounded continuous quantity is isomorphic to the system of positive real numbers. According to Hölder, the range of magnitudes of a quantitative attribute which has additive structure, sustains ratios, and ratios of magnitudes are positive real numbers. For Hölder, operations is the answer to identify whether an attribute qualifies to be a quantitative attribute. But he failed to answer few questions like

1. How you identify two objects as the same in some respect (property) ?

2. What operations between objects determine that a particular attribute of these objects has a additive structure and what is range of magnitudes.

It is interesting from software engineering point of view. Holder axioms help us to find relations between magnitudes for a particular property and to find out whether these properties follow given rules, for example program length or complexity.

3.3 Campbell

Norman Robert Campbell (1919, 1921, and 1928) is known internationally for his important philosophical discussion of the foundations of physical science. Norman Campbell believed that to measure, one must be able to perform a physical operation, a concatenation, such as placing rods end to end to measure length or piling bricks one on top of another to measure weight. He called it physical additivity.

A summary of Campbell’s theory on fundamental and derived measurement: Measurement, according to Campbell, is the assignment of numerals to an attribute according
to scientific laws [7]. The scientific laws derive from the relations demonstrated by an attribute with respect to a certain attribute. The first requirement for measurement is that it must be possible to arrange the attribute to be measured in respect of a given attribute, in an order. The result of this operation is known as an ordinal scale. To do this, it must first be demonstrated by some operations that the relation between the systems is transitive and asymmetrical.

If symbols denote $>$ (greater than) and $\not{>}$ (not greater than) or by the converse $<$ (less than) and $\not{<}$ (not less than)

$A > B$ then $B \not{>} A$ denotes asymmetric \(^1\)

$A > B$ and $B > C$, then $A > C$. denotes transitive \(^2\)

According to Campbell a magnitude must have the relation $=$, this relation is necessary in the construction of an additive scale. He sums up the first conditions for measurement as follows: The first condition of measurement, namely that a magnitude must be capable of order, can now be stated formally as follows. The attribute measured must have transitive and asymmetric relation. Every attribute must be $>$ or $<$ or $=$ every other, and must be $=$ at least one other. The rule for assigning numerals to represent a series in which the above relations have been established is: if $A > B$ then the numeral assigned to $A$ must be greater than the numeral assigned to $B$; conversely, if $B < A$, then the numeral assigned to $B$ must be less than the numeral assigned to $A$. If $A = B$, then the numeral assigned to $A$ must be the same as the numeral assigned to $B$. According to Campbell, the existence of the relation $=$ is one of the things that distinguishes the order characteristic of magnitudes from that which is characteristic of numerals. Numerals, by which is meant simply a group of conventional signs or marks on a piece of paper, obtain
their order by convention. The order is not determined by facts such as the order existing in the family tree. If only one of the many numeral series is used, every member is either greater or less than every other member. There is no relation =. However the most important difference between numerals and magnitudes is that the order of the numerals is conventional while the order of the systems in respect of the magnitude is determined by experimental operations. Numerals have by convention a transitive, asymmetrical relation, now if they are going to be used to represent the order of the systems in respect of a certain magnitude, it must be shown experimentally that the relation between the systems which they represent is also transitive and asymmetrical. If it is impossible to show this then the numerals that have been assigned are meaningless in as much as the conventional relations between them do not express the relations between the systems. There is nothing in the experimentally established relation \( A > B \) that tells what numeral is assigned to \( A \). The rule simply states that it must be greater than that assigned to \( B \). As yet there are no operations to determine by how much \( A > B \), so the assigned numerals cannot reflect a relation that has not been established. In other words, if 2 is assigned to \( B \) and 4 to \( A \), it is impossible to say that \( A \) is twice as great as \( B \) because it has not yet been shown experimentally that \( A \) is twice \( B \). In other words, the numerals can express only those relations that have been shown experimentally to exist between the attributes to which they are assigned.

All physical attributes have this property. Only a few according to Campbell have this structure. They are called fundamental attributes and measurement of them is called fundamental measurement. Ex: mass, volume, pressure, resistance, length. Numbers are
used to represent these measurements because it is easy to find inter relationships between attributes using numerical laws. Campbell has two rules for measuring a physical attribute of an object.

1. They follow an order (<, >, =)
2. Additive law (a + b = c)

According to Campbell, if we can experimentally determine if attributes are fundamental attributes, we can use numerical laws to measure these attributes with respect to fundamental measurement. So for Campbell, fundamental attributes fall into same sense of quantitative structure. So these have internal structure of being quantitative. Fundamental measurement is possible only if there is analogy between physical operations and numerical additions. His physical laws are:

1. The magnitude of the attribute obtained during process of operation in a system must be greater than either body added.
2. The magnitude of outcome from adding bodies A, B, C must depend only on magnitude of those bodies and not on order or method of these additions.

So in his view a clear numerical structure of physical operation is necessary. Once it is clear, we can find interrelationships between attributes using different numeral laws and derived physical laws. But he didn’t agree on logical relationship with attributes which do not have additive structure.

Derived magnitudes (measurements): Physical quantities can be expressed as functions of fundamental measurements. Fundamental physical quantities have perfect physical additive structure analogy to numerical structure. All attributes can’t be expressed in
that form. They can be expressed as functions of Fundamental physical quantity. Ex: density as ratio of mass and volume. According to Campbell, constant is always the measure of magnitude.

Campbell confined measurement to just numerical representation of the operation of physical addition [8]. Major highlights of Campbell theory are: 1. Number representation can be given to system of physical addition. 2. Non additive structure attributes cannot be represented by numerals. 3. Campbell view is to find common property through observation or experimentations. In physical science, he found additivity. It is not necessary that software elements will also have additivity as common factor. 4. Ordinal: His view is ratios of magnitudes of continuous quantity posses all of the formal properties of positive real numbers and this are why numerical concepts (and not just numerals) are implicated in quantitative science.

3.4 Stevens

Stanley smith Stevens was a famous American psychologist. He was famous for book "Handbook of experimental psychology". He introduced levels of measurement and influential in development of operational definitions in experimental psychology.

Problem and issues before Stevens: Campbell is influential in physical sciences with his additive structure and fundamental measurement theory. He argued that empirical relations must have additive structure, properties and characteristics to define the scale. Transitivity, symmetry, reflexivity for classification and transitivity, asymmetry for ordinal. This can be done by experimental operations and scientific investigation. Once
they are discovered independent of numerical assignments, numerals are assigned to objects or events involved to represent this equivalence relations. He divided objects into measurable and non measurable objects. According to him psychology comes under non measurable objects. Different psychologists and mathematicians tried use Campbell measurement theory in psychological sciences but failed to yield results. They categorized representational measurement into external representation and internal representation. Campbell’s ordinal and additive models come under external representation. There is internal representation. Attributes have a putative structure but can be logically defined only when given numerical assignments. It means in extreme case, if numerical assignment not given, attribute may not exit or same component of attribute will be absent. Ex: Assigning numbers to Nationality. Australian 1 French 2 Belgian 3 .Without number, Nationality exists but the ordinal structure doesn’t exist. This is called Internal Representation. Between these representation (external and internal) lies most of psychological attributes. Steven’s Theory of measurement: According to Steven, there exists an isomorphism between properties of numerical series and empirical operations that we can perform with aspects of objects. For Steven, there are no properties or relations logically independent of numerical assignments that exist. Numerical representation can only be of relations defined via operations performed upon these objects. The concept of operations is science’s fundamental theory.

Steven’s Measurement: is defined as the assignment of X to Y according to Z. If \( \rightarrow \) denotes ‘represent’

\[
X \rightarrow \text{numerical values, scores, symbols or abstract system.}
\]
Y → attributes behavior, characteristics, individuals, observations, persons, properties of Experimental units or of objects, responses, situations or things.

Z → particular kind of rule.

Explanation: An operational definition assigns meaning to construct or variable by specifying the activities or operations necessary to measure it or manipulate it. Example: Intelligence (anxiety, achievement) is score of X intelligence test. The test score requires gives measure of intelligence. It is defined by standardized achievement test, test made by teacher or grades assigned by teachers. We have three different operations for same attribute (intelligence). It all depends on time and situation.

Steven’s view of operationism: For any concept, the existence of operation is necessary condition. His understanding of measurement is numerical representation of operationally defined empirical relations. A rule for assigning numerals to objects or events could always be taken as representing empirical relation if those relations were understood operationally defined by rule itself. If operationally definable objects and events don’t exist, then there is no representation (external, internal or ambiguous). For operational point of view, the numbers assigned from operation have to be consistent. At least they have to be nominal scale (classification). All scales (nominal, ordinal, and ratio) follow assigned rules. Like nominal scale is defined for transitivity, symmetry and reflexivity. Ordinal scale like transitivity and asymmetry. Ratio scale empirical relation of definite character.

Steven’s introduction of his own terminology (nominal, ordinal, interval, and ratio) effectively removed the term quantitative from psychologist’s lexicon [8]. He also concluded that scales he used like (ratio, nominal, ordinal, and interval) are par with scales
used in physical sciences. He followed different route which erased quantitative objection and made things simpler for psychology. "If a consistent procedure can be devised for making numerical assignments to objects or events then a measurable attribute is implicated”.

3.5 Suppes

Suppes (1979): Axiomatic- Set - Theory Approach

Suppes supported Campbell theory of measurement. Campbell key is to do experimental work and scientific investigation independent of numerical assignments. The problem before mathematicians is how to investigate systematically. Suppes approach is revolutionary in measurement theory. He provided a program which tells us necessary and sufficient conditions for an empirical structure to be represented in numerical structure. He was within representational theory. Suppes (1950) suggested to philosophers of science to attempt to axiomatize and develop branch of science which involves scientific measurement. A theory is 'axiomatized’ when it is expressed as a set of (ideally, logically, independent) propositions (called axioms). The remainder of theory can be deductible from these axioms. [10]

Axiomatization is set-theoretic when the entities to which theory applies are described as members of a set of some kind and an axiom state condition which this set satisfies.

Since Number System had already been axiomatized set - theoretically, axiomazation of empirical system may bring a connection between (qualitative empirical Systems, Quantitative empirical systems) to Numerical system.

Suppes proposed a four step program for this scientific measurement.
1. A 'qualitative', empirical system is specified as a relational structure that is as a non-empty set of entities of some kind (objects or attributes) together with a finite number of distinct relations between elements of this set. It is required; elements and relations between them are by direct observation. Call them empirical primitives of the system.

2. A set of finite axioms is stated in form of empirical primitives so that they can be directly testable.

3. A numerical structure is identified such that there is a many to one mapping (homomorphism) from the empirical structure to the numerical structure.

4. uniqueness theorem

The specific relation of additivity identified via ratios is between levels of attribute rather than between differences. These two relations of additive are totally distinct and sustain totally different system of ratios and different systems of measurement. Identifying ratio identifies multiplicative laws between quantitative attributes. This fact connects theory of conjoint measurement with what Campbell called Derived measurement. Campbell told constant ratio between mass and volume proves density is derived measurement. Measurement theory didn’t accept it but felt there is some relation of density with mass and volume unidentified. Conjoint Measurement provided logical reason for density. Using that relation, ratio between differences can be identified and measurement can be achieved.
3.6 **Luce and Turkey**

Campbell failed to explain derived measurement in his theory. Luce and Turkey used axiomatic-set-theory approach and gave explanation in conjoint measurement. [8]

The theory of conjoint measurement says that, a combination of attributes in a context enables difference between them relative to joint effects upon third attribute. So for example A tradeoff of increase in two attributes relative to third attribute is done when ratios are equal. Assume Ability and motivation produce performance doesn’t differ between individuals. Persons K and L perform at same level in test. They differ one other in Motivation (M) and Ability (A) but compensate each other in final performance result. K represents Higher level of Motivation than L

L represents Higher level of Ability than K

i.e. $MK - ML = AL - AK$

i.e. the level difference in each of these attributes is traded off against one another to have relative same effect on third attribute ( performance ). If it is possible for one pair of differences, then it is possible for adjoining differences. Suppose there are persons J and H who perform better than K and L.

J represents Ks motivation but more Ability

H represents Ls Ability but more Motivation.

Therefore $MH - MJ = AJ - AH$

So these differences in levels of attribute are traded off for same effective performance.

However $MJ = MK$ and $AH = AL$.

If $MH - MJ = AJ - AH$
Then \( MH - MK = AJ - AL \).

So \( H \) has more motivation than \( K \), \( K \) has more than \( L \) \((MH - MK)\) and \((MK - ML)\) are adjoin. Same \( J \) has more ability than \( L \) and \( L \) has more than \( K \). \((AJ - AL)\) and \((AL - AK)\) are adjoin. If all these attributes are quantitative then two composite differences composed of these adjoin components must be equal.

\[
(MH - MK) + (MK - ML) = (AJ - AL) + (AL - AK)
\]

\[MH - ML = AJ - AK.\]

Additive between differences has been indirectly identified via adjoin component differences. An increase of attribute from \( X \) to \( Y \) not only gives \((Y - X)\) but also gives \(Y/X\). We use \((Y/X)\) factor here. In our example when \( K \) ability is increased to \( L \) ability, there is change in performance and \( L \) motivation is increased to \( K \) motivation, there is change in performance. So they are traded off when the ratios are equal.

\[Ak / AL = ML/MK\]

This test is called Thomsen condition. It tells again about Euclidean axiom equals plus equals gives equals. If equality is replaced by inequality (equal or less than) then condition is know as "Double cancellation". Double cancellation is key in conjoint Measurement. Other important ones are Solvability and Archimedean condition. Solvability Condition: Taking performance \((\text{ability} \times \text{motivation})\) Example: For any difference in ability there exists equal difference in motivation at any point on motivation attribute and vice versa. Archimedean condition: Differences within attribute involved can’t be infinitely small or infinitely large. Relative to other differences i.e. the smaller of two multiplied by natural number must give larger number. Solvability and Archimedean are difficult to solve experimentally. Now let’s implement some conjoint measurement to other relations. Let’s
take ratios. If there is one relation of addition in a continuous quantity then there are finite numbers of relations. To find one relation is important. One relation of addition is identified by differences. Conjoint measurement says that trade off is done in increase of attribute if specific difference within attributes are same relative to third attribute. The specific relation of additive identified within each attribute is connecting these differences.

Conjoint Measurement: Logic for straight line is like this. AC has two discrete adjoining intervals AB and BC. \( A' \ C' \) has two discrete adjoin intervals \( A' \ B' \) and \( B' \ C' \). 

\[ AC = A' \ C' \]

If \( AB = A' \ B' \) and \( BC = B' \ C' \).

In Context of discrete adjoining Intervals on a straight line equals plus equal gives equals. If is true for straight line, (for difference in intervals) it will be true for (difference in attributes) Conjoint Measurement generalizes this idea to combination of attributes in a context which enables difference within two attributes to be matched between them relative to joint effect upon a third attribute. Suppose there are attributes \( X, Y \) related to attribute \( Z \). \( Z \) increases only with increase in \( X \) and \( Y \). \( Z \) has same effect by increasing specific amount on \( X \) or by increasing specific amount of \( Y \) (difference in adjoining values). Then the difference in increase of values of \( X \) and \( Y \) are equal.

Generalization is if two discrete but adjoining differences within attribute \( X \) can be matched with two discrete but adjoining differences within attribute \( Y \) and if these attribute are quantitative then \( X \) intervals = \( Y \) intervals w.r.t. \( Z \) Example: Performance on a test is influenced by motivation factor and ability factor and experimental test shows
difference in these factors shows different performance, and if different levels of motivation, ability can be identified then conjoint measurement can be applied on this test.

Advantages of Conjoint Measurement: [8] Axiomatic set theory approach to measurement provided a methodological basis for Luce and Tukey to develop Conjoint Measurement. Extensive Measurement relies on a concatenation operation. The effects of the operation depend almost entirely on a single attribute. But every object is complex in the sense that it has indefinite properties and has indefinite relations with one another. Any operation of bringing two objects together in some way with an operation will have effects and not necessity on only one attribute. They are chances. It may happen for one attribute for example length of rod. But normally operation has effect on other attributes. Our capacity to identify concatenation operation suitable for extensive measurement of attributes that interest us depends on existence of very special causal relations, as well as sensory-motor capacities and how we observe these attributes. "As nature loves to hide", it’s difficult to observe these attributes in psychological measurement but relatively smooth in physical measurement. Campbell’s derived measurement concept proved that there are other quantitative attributes than fundamental attributes (additive structure). Campbell’s theory of constant in ratio for derived measurement may be wrong but it shows they are other ways to find it. Density and volume trade off against one other relative to mass. The relation of additivity for this trade off is ratio. Example: Take brick of gold and aluminum of same mass. Increase in volume between Aluminum and Gold equals increase in density between Aluminum and Gold. $\frac{VA}{VG} = \frac{DG}{DA}$ This means the relationship between density and volume is not arbitrary but it is testable and scientific hypothesis. This same logic applies to all instance of derived measurement in
physics. So psychology must renew the research in quantitative psychology attributes who were discouraged before as Campbell told quantitative attributes are fundamental attributes or derivatives of fundamental attributes.

Misuse or Disadvantages of conjoint measurement:

1. The necessary condition in conjoint measurement is attributes have to be quantitative. But in most of cases they are assumed.

2. Performance = motivation * ability. Here motivation and ability are quantitative attributes for this performance. But we need to clearly distinguish them as quantitative attributes outside this test. They have to be quantitative attributes independent of performance test. Conjoint measurement needs attributes which are hypothesized and experimentally investigated about there quantitative structure.

3. Conjoint measurement needs quantification. Conjoint measurement is one conceptual source in that fills a specific debilitating gap in quantitative psychologist methodological array. It indicates place to start.

4. Conjoint measurement gives relation between attributes in quantitative concept. It is good concept in quantitative psychology. Suppes and Zinnes axiomatic set-theory approach tells more about empirical structure investigation and then representing them with numerical structure. It is surface-level theory. Measurement in psychology has to look for surface level theory with conjoint measurement.
CHAPTER 4

Fundamental Problems of Representational Measurement

4.1 Representation Theory

Representational theory is understanding of nature through empirical observation or experimental research and provide some symbols to measure them.

Permissible transformations: [13] Relationship has to be preserved when an attribute measured in one scale is transferred to other scale of measurement. For example the distance between two cities is measured in miles. If distance is measured in kilometers, value of distance will be multiplied but relation doesn’t alter. So we can say distance has permissible transformation between different units

Levels of measurement: There are different levels of measurement that involve different properties (relations and operations) of the numbers or symbols that constitute the measurements. Associated with each level of measurement is a set of permissible transformations. The most commonly discussed levels of measurement are as follows:

Nominal: We can also say classification. Two objects are grouped in one classification if they have same value of attribute. Ex: Americans, Indians, Africans. Permissible transformation is one to one mapping.

Ordinal: As name suggests, objects are ordered with order relation that attribute has on each other. Two things x and y with attribute values a(x) and a(y) are assigned numbers m(x) and m(y) such that if m(x) > m(y), then a(x) > a(y). Permissible transformations are any monotone increasing transformation, although a transformation that
is not strictly increasing loses information. Ex: Grading, rating.

**Interval:** Measurement is done with difference between values of an attribute. If \( m(x) - m(y) > m(u) - m(v) \), then \( a(x) - a(y) > a(u) - a(v) \). Ex: Temperature scales like Fahrenheit and Celsius.

**Log-interval:** Attributes of an object are represented as ratio of numbers. If \( \frac{m(x)}{m(y)} > \frac{m(u)}{m(v)} \), then \( \frac{a(x)}{a(y)} > \frac{a(u)}{a(v)} \). Ex: Density (mass/volume).

**Ratio:** The difference and ratio between attributes are represented with difference and ratio between numbers. Ex: Mass, length.

### 4.2 Fundamental Problems of Representational Measurement Existence

The most fundamental problem is building representation. We have observed an empirical structure for an attribute or property of an object. What we generally do his to find similar pattern in numerical structure. If it is existing in numerical structure, we determine the measurement. Then using these numerical structures and computational methods we derive information from empirical structure. The problem is it is limited to scope of our observation. The other problem is uniqueness. We are assigning same numbers to represent different properties of objects. It can create confusing. Temperature of Kent is 30 is not same Kent is 30 miles from Akron. One is interval scale and other is ratio scale. What we need to do is formally define scale of measurement.
What is software measurement? It is not the same sense of measurement as we do in physical sciences or classical measurement as defined in software literature. A clear formal definition of software measurement is necessary to know what we are doing in software measurement.

In literature, software measurement is used in various areas of software development. Software is a combination of physical needs i.e., hardware and psychological needs i.e., human interaction. It depends on requirements, design methodologies, implementation, hardware limitation, cost effect, human resources, and mental conditions. All questions we ask in these areas are software issues and need software measurement. Software measurement cannot be explained as a single theory. It means we have to subdivide these areas and develop before meaningful theories are available.

In this chapter, we discuss some of software measurement issues with classical measurement issues. Why do we want to do it if you believe software measurement is not as classical measurement? This question leads me to explain about Halstead Maurice and his theory on software science elements.

The first person to systematically do what we do not consider software measurement was Maurice Halstead. Halstead was a professor at Purdue University. His software measurement work was called software science. Halstead is not aware of Measurement
theory. The question may rise why Halstead’s work should be considered with measurement theory as we don’t consider software measurement a part of classical measurement. Although Halstead work is not classical measurement, some of its ideas resembled classical measurement. In early 1990s many researchers were using measurement theory to try to understand software measurement.

In this chapter, we compare software measurement, at least, from the limited scope of software science, with classical measurement theory. Our purpose is to use classical measurement to help us understand issues in software measurement so that once software measurement is better understood we may be able to develop a theory for at least part of software measurement. Our comparison of software measurement and classical measurement will be done though an extended imaginary conversation between Halstead and Campbell.

Our simple setting includes just Halstead and Campbell. Halstead is in the process of finalizing his software science ideas, and Campbell has fully developed his ideas on measurement, which means measurement in the physical sciences. The assumptions are that Halstead thinks that his work is a type of measurement or is, at least, related to classical measurement, and he wants to discuss them with Campbell.

**Halstead:** I want to better understand software and what happens when someone writes software programs.
**Campbell:** It sounds like you want to measure software. When one thinks about measuring something, then one needs to clearly identify the things or objects being measured and the attribute or characteristic of interest.

**Halstead:** Does one then measure the object or the attribute?

**Campbell:** Good question. In fact, when one measures an object, one is really measuring the amount of the attribute in that object. For example, when I measure my own weight, I’m not measuring all of me, but I’m simply measuring the ”amount” of pounds or kilograms in me.

**Halstead:** That makes sense.

**Campbell:** An important point is that one must be very specific regarding what one wants to measure. What is it that you want to measure with respect to software?

**Halstead:** I want to measure length of a software program. I found it as property of a software program. I found that program length varies when same algorithm is written in different programming languages.

**Campbell:** nice. What do you want to measure this property?

**Halstead:** I need to find the relations program length has on a software program. It
gives me relation between program logic and program length. It also helps to compare complexity in writing different programs.

**Campbell:** It is reasonable. How did you measure this program length? Did you used any countable attributes or say fundamental measurements. For example when I want to measure density of particular object, I didn’t find appropriate direct metric which suits its property. I measured it using other direct metrics like mass and volume.

**Halstead:** I have two fundamental measurements $\eta_1$ and $\eta_2$. $\eta_1$ counts distinct operands in program and $\eta_2$ counts distinct operators in program. Operands can be variables, constants. Operators can be conditional statements, loop statements, jump conditions, end of statements, mathematical operations. Sum of $\eta_1$ and $\eta_2$ provide unique token for the program. I call this vocabulary of program ($\eta$). The program length can be derived from sum of total number of operands and operators in a program.

**Campbell:** What is relation of program length and operators and operands? For example to measure force applied on an object we use mass of object and velocity through which it is traveling. The force increase if mass of object is increased. They have linear relation. It is found through empirical observation and proved through experiments and laws supporting this theory.

**Halstead:** Program length has linear relation with token or vocabulary or logic of a program. If program logic is modified, for example program divided in sub programs
or new features added to same program, its program length changes. It is proved by
empirical observation by counting operands and operators of different programs. Exper-
imentially it is proved by taking different 13 large, small programs and deriving using
\( \eta_1 \log \eta_1 + \eta_2 \log \eta_2 \). The results show 1 percent of error difference from empirical
observation values.

**Campbell:** Your empirical observation is right and provided us a starting point. But
it has to be proved by some universal laws. These laws must prove that vocabulary or
tokens of program has direct relation with program length. I agree that software pro-
gram depends on programming language it is written in. But it also depends on mental
conditions of a programmer, time factor and cost effect. The basic values \( \text{eta1} \) and \( \text{eta2} \)
does not provide us reasons for these factors.

**Campbell:** What are the uses of program length? What will a programmer or de-
signer will achieve knowing the value of program length.

**Halstead:** It provides complexity of given program in given programming language
for a given algorithm. It may help to compare with other programs. The lines of code
written in a program could influence complexity of a program.

**Campbell:** Software is not a single unit. It is divided into many subcategories like
specifications, design, implementation, maintenance and testing. Programs can be se-
quential or modular. Property depends on internal elements and external elements. You
need to give a formal definition of complexity with respect to sub category it is measuring.

Halstead: Could I just define complexity to something in lines of \( \eta_1, \eta_2, N_1, N_2 \).

Campbell: The answer is yes and no. Yes it becomes a defined measurement. No because you are just giving name to equation.

Campbell: Software is not a single unit. It is divided into many subcategories like specifications, design, implementation, maintenance and testing. Programs can be sequential or modular. Property depends on internal elements and external elements. You need to give a formal definition of complexity with respect to sub category it is measuring.

Campbell: Lets us discuss few more basic things about an attribute. Magnitudes of an attribute exhibit few properties if they are in ordered structure. They are transitive and asymmetric properties. If are three magnitudes \( a, b, c \) then if \( a > b \) and \( b > c \) then \( a > c \). It is called transitive. If \( a > b \) and \( b \npreceq a \) then it is called asymmetric.

Other property they need to exhibit is additivity. It proves how much other magnitude is greater or lesser than other magnitude. It also helps to compare same attribute in two objects. Additive can defined liked \( a + b = c \).

Other property is commutative. Magnitudes of attributes must provide same result with change in order of operation. The rule is \( a + b = b + a \).
Other property is associative. Grouping operations of magnitudes of an attribute must provide same result. The rule is \((a + b) + c = a + (b + c)\).

So can prove that program length metrics exhibit these properties?

**Halstead:** Program length metric contain two fundamental measurements \(\eta_1\) and \(\eta_2\). Let say it as \(\eta\) (vocabulary) i.e. sum of distinct operators and operands. \(\eta\) follows transitive and asymmetric properties. Three distinct programs have different vocabulary and it follows transitive and asymmetric properties. It also follows additivity. When two distinct programs are added they have distinct vocabulary and result program have sum of these vocabulary. \(\eta\) (vocabulary) also follows commutative and associative as order of operation and group operation does not change results.

**Halstead:** Program length \(N\) is sum of total operands and operators in a program. It follows transitive and asymmetric properties. But it also has symmetric property. Two different programs can have same program length. Program lengths do not follow additive structure. When two programs are added there is possibility that they have same operands or operators. The result program length can be less than sum of individual programs. Two programs are added because requirements have changed and it may lead to change in design. Change in design may be change in implementation details and more lines of code. It also changes understandability and maintenance. Program length follows commutative rule. The orders of operation do not change program length when two distinct programs are added. Program length do not follow associative rule. Two
distinct programs a and b may have common vocabularies but b and c may not have common vocabulary. Here program also depends on different factors of requirements, design, implementation and maintenance.

**Campbell:** That makes sense. What I understood from above dialogue is properties followed by attribute in physical science are followed in few cases and contradictory in other cases. The reasons of contradiction are reasonable because of different structure and different layers a program is made of. I feel software metrics may not be fully explained on measurement theories. New theories have to develop. Each stage of software engineering has to treat separately. You can define properties in a single theory.
CHAPTER 6

A Theory for Software Measurement

As we have seen in this thesis, there is a well developed theory for classical measurement, the type of measurement often seen in physical sciences. However, there is no such theory for software measurement. Of course, one may claim that there is no need to be a theory for software measurement. However, as we stated earlier in this thesis, if there is no theory for software measurement, how can one know if one is doing software measurement, or how can one know if one has or has not defined a software metric? It may not be the case that an extensive software measurement theory needs to be developed, but certainly, at least, a clearly defined basic theory needs to exist. Please note that we are not saying that an extensive theory should not be developed; we are merely saying that it should not be the case that no software measurement theory exists.

Further, we believe, and we have already hinted at this belief and we will say more about it in this chapter that software measurement is too diverse an area in that there are too many different types of measurement-like activities going on in software measurement for a single (useful) theory to exist. Thus, we believe that software measurement will need to be divided into sub areas, and then the software measurement community will need to develop measurement theories for the various sub areas.

In this chapter, we discuss a couple ideas regarding developing theories for software measurement, and then we discuss two critical, in our opinion, problems with software
measurement. The two problems are complexity and binary operations on software entities. Thinking of dividing software measurement into disjoint ”pieces” and then defining a theory for each piece, we have three ideas to do it.

The first idea is if we could make a list of possible properties of software metrics, and then we could classify metrics based on subset of properties which each satisfies. Here is a small example. We listed properties an attribute has to posses from classical measurement point of view, properties a software attribute need to posses from point view of people working in software metrics like Fenton, Basil, and Weyuker. Properties of two software attributes like program length and program effort or complexity.

**Properties a measurement has to posses according to measurement theory:**

1. An attribute must have a range of magnitudes.

2. Magnitudes must have relation.

3. Order of magnitudes is important i.e. Transitive and Asymmetric.

4. Monotonic $a < a + b$

5. Order of operation doesn’t matter. $a + b = b + a$

6. Additive structure. $A + B = C$

7. Continuity. If there exist a as unit then there exist b where $b = ra$ where $r$ is some real number.

8. Has to posses some level of measurement.( ratio , nominal, interval, absolute)

**Properties a software measurement need to posses:**
1. Non negativity

2. Zero (null)

3. Level of measurement


5. Additivity. \( a + b \leq (a + b) \)

6. Order of operation (Commutative property)

7. Too coarse. Metric values obtained must provide clear difference.

8. To sensitive. Metric values must not be too sensitive.

9. Different programs can have same measurement value.

10. Two same kinds of programs can have different measurement value.

11. \( a + (b + c) \) equal to \( (a + b) + c \) (associative property).

**Program Length**: was defined as size of a program. Program is group of modules.

Each module contains statements. Properties program length posses:

1. Non negativity: Program length can't be nonnegative. \( N \geq 0 \)

2. Zero (null): If program have no statements. It can be null or zero. \( N = 0 \)

3. Monotonic: If program statements are added to program, program length has to greater than initial program.

4. Weak order : program lengths has to follow relation order .
5. Weak order of associatively: program length has to same independent of order programs are associated.

6. Archimedean order: if \( c > d \) then \( a \circ n c > b \circ n d \). It means the relation between program length \( C \) and \( D \) must not change on some operation with Program A and B. Here ‘\( \circ \)’ is some binary operation.

7. absolute scale.

**Programming Effort (Complexity):** it defines amount of time taken to complete the program. It determines amount of program lines of code in given amount of time.

**Properties programming effort posses:**

1. Non negativity:

2. Zero (null): If program have no statements. It can be null or zero. \( N=0 \)

3. Modular interaction: If Modules are independent, then there will be less complexity.

4. Specification: If program or module is more clearly specified then it takes less programming effort.

5. Lines of code: More lines of code then more program effort.

6. Monotonic: If modules with no interaction are added to a program, program effort of program is no less than sum of programming effort of individual programs.

\[
PE(P) \geq PE(P1) + PE(m1).
\]

7. Archimedean order : if \( c > d \) then \( a \circ n c > b \circ n d \)
8. Interval scale.

Here in above properties program length and program effort match in non negativity, zero, monotonic and Archimedean order. Both of them did not agree on additive property. The properties of program effort as other properties like modular interaction, specification. The other basic thing to take care is level of measurement. You need to identify these properties also. Measurement may be produced by Mean, median or mode. Mode provides us nominal scale. Median provides us ordinal scale and mean provide interval scale. You cannot misplace them and it makes no sense in measurement. In our example both of these metrics have different scale. It looks like both have major different properties. So as of now it makes us to classify them into two. It is a small example, we need to further develop this by collecting all properties of different software attributes and classifying them. Further, we could establish what could be done with each equivalence class of metrics. If such a classification scheme were implemented, it could be useful in that we would then know what properties our metrics have and how the metrics and the values or measurements gotten from them could be used.

The second idea is a modified but similar procedure of gathering all properties of different software metrics. Again, prepare a list of software metrics’ properties. However, this time group the properties based on criteria such as: similar properties, related properties, or other meaningful connections among the properties. For each collection of properties, establish results for the metrics which have these properties.

The third idea begins from the point of view of software engineering. Classify the activities and products of software engineering into similar groups based on properties, uses
etc. For each of these groups consider the metrics that are relevant, and establish the properties and uses that each group’s metrics have.

The second part of this chapter is about complexity and binary operations. In this paragraph, you can also take about Campbell’s defined measurement or metric. This is a metric that is defined by writing a mathematical equation but without an understanding of what the equation describes. Complexity in software measurement shows how bad defined measurement can be. Really all we have for complexity are vague definitions and many defined metrics.

The other problem is binary operations. Binary operation is defined as $m(A + B) = m(A) + m(B)$. Here ‘+’ is binary operation. One of important property in measurement is additivity. Additivity is proved through some binary operation. Let us take Halstead program length (N). Additivity is proved for program length (N) by adding two programs. But additivity fails for unique operators and operands ($\eta$). So for some particular notions binary operation works. If it changes additivity does not work. Sometimes by changing binary operation we get additivity. The example is $\eta$. If we change our operation as $\eta_1^{n_1} + \eta_2^{n_2}$, it provides us additivity. So now it tells us binary operation is a key to get this property. Let us take Weyuker properties. Assume programs P and Q with complexity $|P|$ and $|Q|$ and have equal complexity $|P| = |Q|$. Now if we combine program R with programs P and Q, there complexities changes and are not equal. So additivity does not work. So this particular binary operation does not work in this case. May be it is sensitive in this case. May be we need to use different operation to get additive property. It is the case that there may be multiple possible binary relations
associated with a given metric, and with different relations, a metric may have different properties, including being or not being additive.”
CHAPTER 7

Conclusion

A number of changes occurred in measurement. Measurement in psychology did not produce good results when they approached representational theory of physical measurement. Steven theory, set-axiomatic theory and conjoint measurement provided them better approach to follow. We have similar cases in software engineering measurement where software metric like Halstead software elements followed representational theory of physical measurement. The problem here is they didn’t explain definition of property they are measuring. A different approach is necessary in software metrics to get accurate results.
BIBLIOGRAPHY


[14] “Measurement theory principles in software measurement”, Austin Melton