Software Metrics & Software Metrology

Alain Abran

Chapter 8
Function Points: Analysis of their Design

Agenda

This chapter covers:

- The origin of Functional Size Measurement
- The global design of Function Points - FP
- Analysis of the design of Unadjusted Function Points
- Analysis of the design of the Value Adjustment Factor
- What is a Function Point?
- What if the Function Points weights were dropped?

Agenda

This chapter covers:

- **The origin of Functional Size Measurement**
- The global design of Function Points - FP
- Analysis of the design of Unadjusted Function Points
- Analysis of the design of the Value Adjustment Factor
- What is a Function Point?
- What if the Function Points weights were dropped?

Introduction

- A very large number of software measures have been proposed in past decades, but most are not yet close to international standardization.
- By contrast, the measurement of the functional size of software has a unique status in software engineering: it is the first, and only, type of software measurement with international standards adopted by the ISO.
  - However, just because a measurement method is recognized as an international standard does not guarantee that its design is perfect.
  - In this chapter, we assess one of the functional size measurement (FSM) methods, Function Points (FP), pointing out some of its weaknesses.

Introduction

- It is particularly important to conduct this assessment, since the basic design of Function Points has subsequently been reused in the design of a number of variants of the method.
  - In general, the weaknesses inherited by these variants were not corrected
    - see chapter 9 for a discussion on the design of Use Case Points.
  - the other hand, some of the weaknesses that we identify were corrected in the design of COSMIC Function Points – see chapters 11 and 12.

- The analysis method presented in this chapter can also be used for analyzing other measurement methods:
  - in particular those with ‘weights’ or ‘points’ in their designs.

The Origin of Software Functional Size Measurement

- The technical size of software is used to measure software products from a developer's perspective:
  - number of lines of code, number of components, number of modules, number of sub-systems, etc.
  - It is typically based on the counts of some entities, and can be used in efficiency analysis to improve, for example, design performance.
- The functional size of software is used to measure software products from a user's perspective.
  - It must be independent of technical development and implementation decisions, and, as such it can be used to compare the productivity of different techniques and technologies.

Box 1: Albrecht Key Contribution

At a time when software size meant 'lines of code' for the majority of software engineers, many of whom wrestled with problems where lines of code meant different things, depending on the programming language, Albrecht proposed a method of sizing software which was independent of the programming language used.

The Origin of Software Functional Size Measurement

- In 1984, the International Function Point Users Group (IFPUG) was formed to foster and promote the evolution of the Function Points method.
  - Much work went into the subsequent releases of the method to include rules allowing an interpretation of functionality which was increasingly independent of the physical implementation of the software.
    - The major contribution of IFPUG to the FSM field has been to document measurement rules aimed at improving the level of uniformity in the application of this measurement method.
    - The base structure of Function Points, though, has remained unchanged from what was proposed by Albrecht in 1979 & 1984.

Agenda

This chapter covers:

- The origin of Functional Size Measurement
- **The global design of Function Points - FP**
- Analysis of the design of Unadjusted Function Points
- Analysis of the design of the Value Adjustment Factor
- What is a Function Point?
- What if the Function Points weights were dropped?

The Design of Function Points - FP

- **Global design**
  - In the IFPUG Counting Practices Manual, Function Points – FP - are described in terms of:
    - definitions,
    - rules,
    - decisions tables,
    - an interpretation guide, and
    - examples.
  - At the highest level of abstraction, there are 3 main parts to the measurement process of Function Points (Next Figure):
    - the unadjusted functional size (UFP),
    - the Value Adjustment Factor (VAF), and
    - their combination: The Adjusted Function Points - AFP

The Design of Function Points - FP

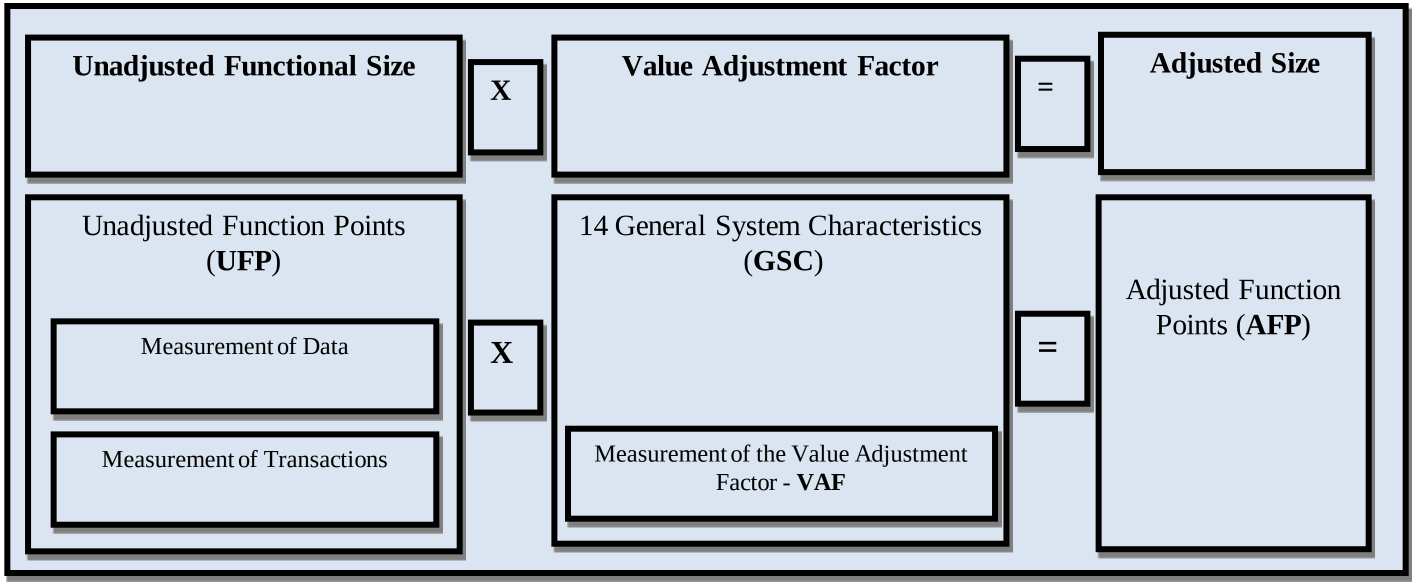

Figure 1: Global model of Function Points – Albrecht-IFPUG

The Design of Function Points - FP

A. The unadjusted functional size is calculated from the counts of all the individual functions of the software application being measured.

- This 1st measurement result represents the addition of the parts of the whole, and is referred to the *Unadjusted Function Points (UFP).*
- The process of measuring the UFP is broken down into 2 steps:
  - The measurement process for the Data (i.e. logical data files),
  - The measurement process for the Transactions.

B. The value adjustment factor (VAF) is calculated by adding the values attributed to the 14 General System Characteristics (GSC) of the software application as a whole.

C. The final adjusted functional size is then obtained by multiplying the 2 figures together, that is, UFP times VAF

- \- referred to as the *Adjusted Function Points (AFP)*.

The Design of Function Points - FP

- The measurement process for the **data** consists of 5 steps, referred to as Fi in next Figure and Table.
  - In the Figure, the inputs of each step are shown in the left-hand column, and the outputs are shown in the right-hand column.
  - If an output of step Fi is reused as an input to step F(i+1), this is indicated by an arrow towards the following step F(i+1).

The Design of Function Points - FP

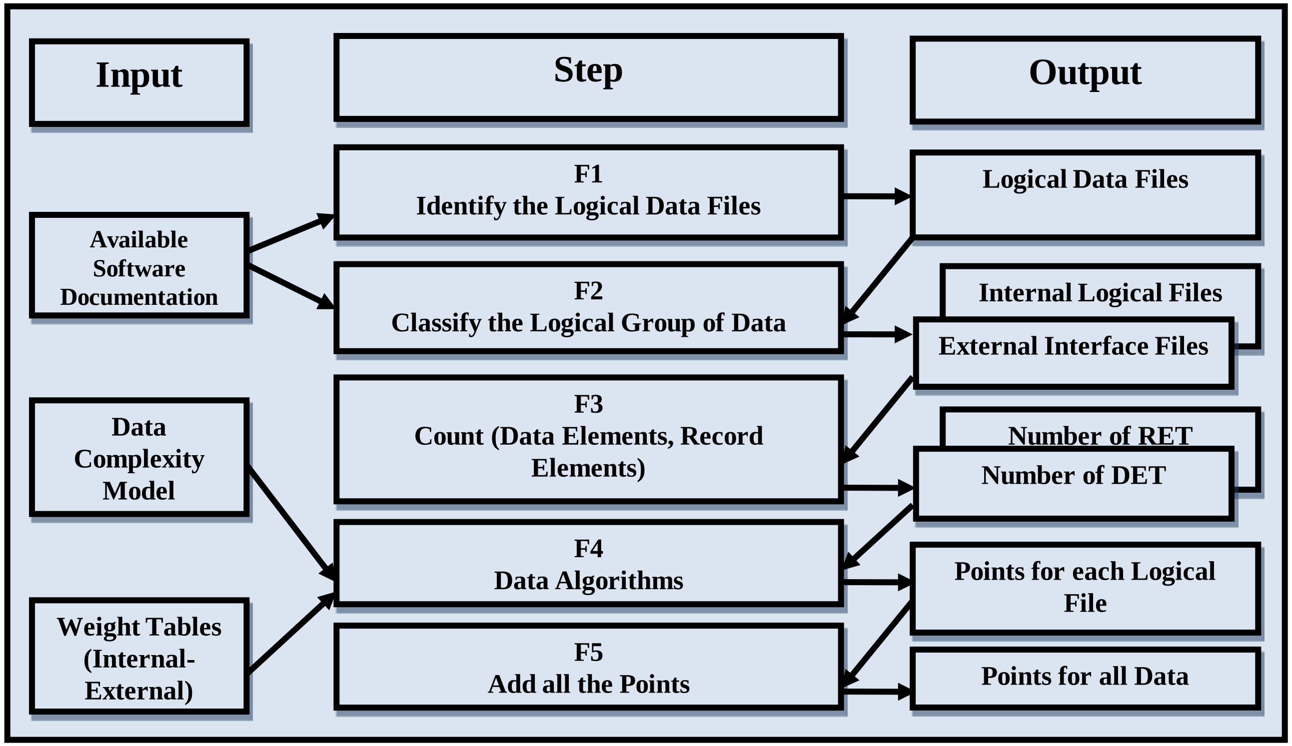

Figure: Data Measurement Model

The Design of Function Points - FP

Table: Data Measurement Steps

ID	Step Description
F1	The input in this step is the available information about the functions of the software (from the software documentation or from the software itself). From this information, the measurer identifies the logical data files.
F2	The same available information is examined to determine the boundaries between the application (or the project) being measured and the other external applications, in order to classify the logical files as internal logical files (e.g. files created or modified within the software being measured) or as external interface files (e.g. files that are read by the software being measured, but not modified by this software).
F3	The files identified in F2 are used as input by the measurer to identify and count, for each of these files, the number of data element types (DET) in each file, as well as the number of record element types (RET) for the internal logical files and the file types referenced (FTR) (e.g. accessed) by each of the external interface files.
F4	The measurer applies the Function Points data algorithm using five types of input: – the lists of internal logical files and external interface files, – the count of their record elements (RET), – the count of their data element types (DET), – the complexity model for data (i.e. the structure of the complexity table with its two dimensional axes), and – the weights within the tables (i.e. the points) for the files (internal and external) which correspond to the step-function.
F5	The points from each file are added together, leading to the number of UFP for the data (that is, for all the internal logical files and all the external interface files).

Analysis of the design of the measurement of the data in Function Points in Abran & Robillard [1994].

The Design of Function Points - FP

- This measurement process for the data (logical files) calls upon several implicit models [1], such as models of:
  - functional specifications,
  - boundaries between applications, and structures,
  - as well as complexity and weights for the logical files.

    - It must be noted that none of the relations of these models is based on any experimentally justified theories in any precise framework, either in Albrecht's original paper or in the IFPUG documentation:
    - they are still only described in a set of rules established by the normative committee of IFPUG, the Counting Practices Committee.

[1] An implicit model is a set of intuitive relations between different objects or concepts.

The Design of Function Points - FP

- ***Measurement of the Value Adjustment Factor***
  - The adjustment process of the Function Points measurement method is intended to adjust the unadjusted functional size by multiplying it by an adjustment factor that reflects the complexity of the processing of the software and of its development environment.
    - Next Figure shows a model of the measurement process that determines the VAF, while Table 2 presents the corresponding detailed measurement steps.

The Design of Function Points - FP

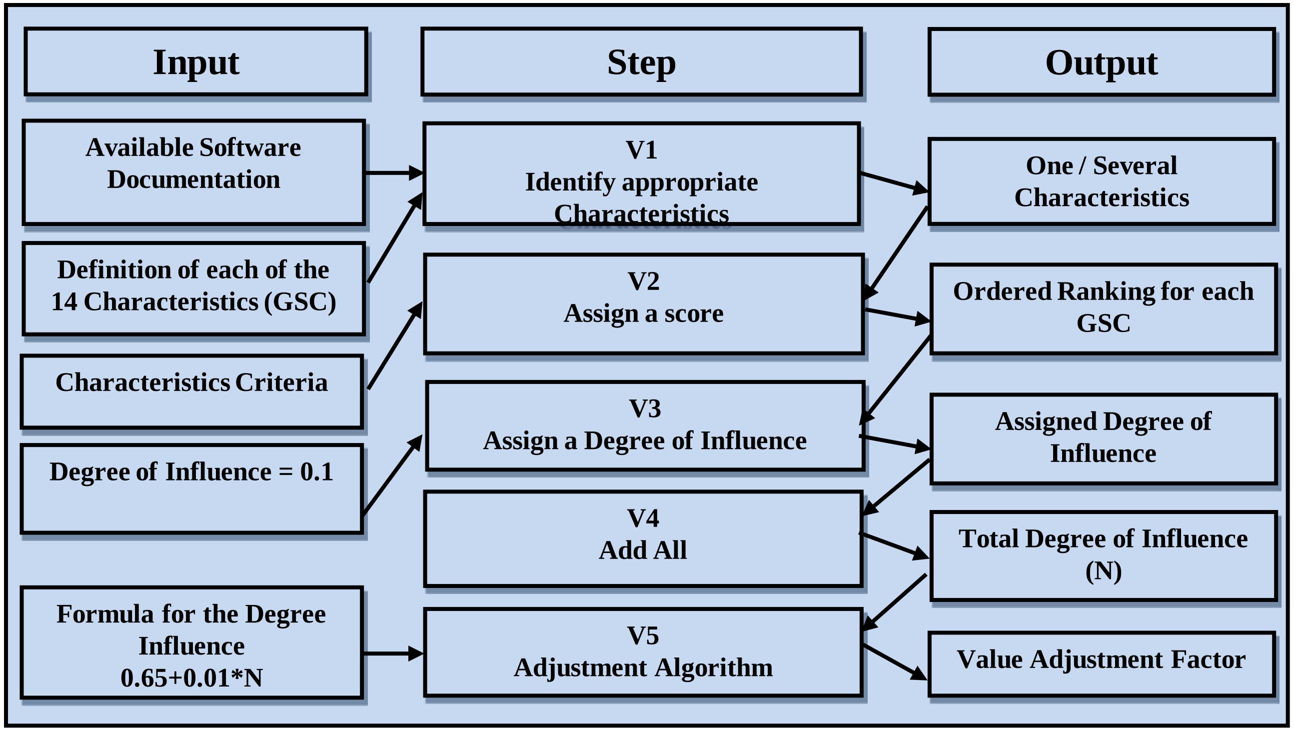

Figure: Model of the Value Adjustment Factor

The Design of Function Points - FP

- Using Albrecht's design, this adjustment factor will vary from a minimum of 0.65 (when all 14 characteristics have all been assigned a so-called 'degree of influence' of 0) and a maximum of 1.35 (when all 14 characteristics have been assigned the maximum degree of influence of 5).

  - This linear transformation provides a maximum adjustment of +/- 35 % to the functional size.

Weaknesses of the Function Points Measurement Design

ID.	Step Description
V1	The available information about the software, as well as the definitions of each of the 14 GSC of the application, are taken as inputs to identify which of these 14 characteristics are present for this software.
V2	Each of the characteristics present in the software being measured is analyzed using its corresponding category criteria to classify it into an ordered ranking from 0 to 5, each step representing 5 distinct irregular intervals.
V3	Each of the numbered intervals assigned for any characteristic is multiplied by a value of 0.1, which gives a degree of influence to each: the range of the derived quantities is then between 0.0 and 0.5.
V4	The degrees of influence assigned to each characteristic are added together to obtain a *total number* of degrees of influence (say, N) – min = 0.00 and max = 0.70.
V5	The following linear transformation is used to obtain the total number of degrees of influence for the whole software being measured: $VAF = 0.65 + 0.01 * N$

Table: Value Adjustment Factor: Measurement Steps

Agenda

This chapter covers:

- The origin of Functional Size Measurement
- The global design of Function Points - FP
- **Analysis of the design of Unadjusted Function Points**
- Analysis of the design of the Value Adjustment Factor
- What is a Function Point?
- What if the Function Points weights were dropped?

Weaknesses of the Function Points Measurement Design

- School students learn the rules that govern the manipulation of quantities very early on. These rules are referred to as the admissible mathematical operations and are dependent on scale types. For example:
  - on a **ratio** scale, quantities can be **added** and **multiplied**,
  - on an **ordinal** scale, quantities **can only be ordered**.
- How well doe Function Points fare on these concepts?

- ***Properties of measurement scales***
  - There are 5 scale types: nominal, ordinal, interval, ratio, and absolute – see Table 3 (next Slide).
  - See also the admissible mathematical operations (i.e. equality, addition, multiplication, division) permitted for each type of scale – see Table 4 (next Slide).

Weaknesses of the Function Points Measurement Design

Scale type	**Description** (See also section 3.5, Chapter 5)
Nominal	This scale type is used to name objects or events. It is used only in identifying requirements, and the only quantitative notion associated with it is that of equality. Only non-parametric statistics can be used.
Ordinal	This scale type is used to put objects in order, based on a criterion that may be subjective, but it is preferably objective. Ranking order statistics can be used, as well as those that apply to the nominal scale.
Interval	This scale type (also called the cardinal scale) is used to determine the difference between ranks. It is a continuum between two points which are not necessarily fixed. With this scale, objects can be distinguished and ranked; moreover, the differences between the ranks can be measured. The mathematical average can be used, as well as all the statistical methods that apply to the ordinal scale.
Ratio	In this scale type, it is significant to multiply a measurement by a non-negative value. It is then possible to say that an item X has *n* times the value of item Y with respect to a given attribute. It follows from this that the value zero has a special significance for that attribute. This allows us to distinguish a ratio scale from an interval scale. The calculation of percentages is allowed, as well as all the statistical methods that apply to the interval scale.
Absolute	The absolute scale type possesses a unique origin from which to start the measurement process. This allows us to count entities, and there is only one way to do this (Fenton, 1991). Here, only the *identity* transformation *(f(x)=x)* is admissible. All the statistical methods for the previous types of scale apply.

Table 3: Properties of Scale Types

Weaknesses of the Function Points Measurement Design

Scale type		Admissible Transformation	Operations	Examples
Nominal	(R,=)	f unique	Name, distinguish	Colors, shapes
Ordinal	(R,>=)	f strictly increasing monotonic function	Rank, Order	Preference, hardness
Interval	(R,>=,+)	f(x)= ax + b, a>0	Add	Calendar time, temperature (degrees Celsius)
Ratio	(R,>=,+)	f(x)= ax, a>0	Add, multiply, divide	Mass, distance, absolute temperature (degrees Kelvin)
Absolute	(R,>=,+)	f(x)= x	Add, multiply, divide	Entity count

Table 4 : Measurement scale types and admissible transformations

Weaknesses of the Function Points Measurement Design

- **Analysis of the measurement design for the Data**
  - This section identifies the measurement scale types within the measurement design for the data (Figure 2) and analyzes the corresponding mathematical operations.
  - The numerical assignment of size points is described in 3 steps:
    - a logical file is first classified as simple, medium, or complex,
    - a weight (in points) is assigned depending on its position in a reference table, and, finally,
    - the points are added together.

Weaknesses of the Function Points Measurement Design

Box 2: Example of the measurement of data in IFPUG Function Points

A software application has three internal logical files as follows:
– file 1 has 1 RET and 5 DET,
– file 2 has 3 RET and 21 DET,
– file 3 has 6 RET and 26 DET.

Using the table of weights for the Internal Logical Files:
– file 1 is classified as simple with a weight of 7,
– file 2 is classified as medium with a weight of 10,
– file 3 is classified as complex with a weight of 15.

Adding their weights yields a total of 32 FP for these three files.

Weaknesses of the Function Points Measurement Design

- In the measurement of the internal and external files, 3 different types of attributes are taken into account in the measurement design:
    1. the data element types (DET),
    2. the record element types (RET), and
    3. the file types referenced (FTR).

- It is to be noted that these 3 types of attributes are not independent, but organized in a hierarchical structure:
    - a record consists of 1 or more data element types (DET), and
    - a logical file (FTR) consists of 1 or more record types (RET).

Weaknesses of the Function Points Measurement Design

- These represent different levels of abstraction of the data in software:
  - the RET represents a structure of DET, and
  - the FTR, a structure of RET.
- To identify the types of scale and to analyze their use in this measurement design, the numerical assignment rules are decomposed for the internal logical files as follows:
  1. A file is analyzed, and then each of its DET and RET or FTR is identified and counted; this corresponds to the addition of numbers in an <u>absolute scale type</u>.
  2. Each of these DET and RET counts is compared to the 3 DET and RET intervals within the 3 ranks listed in Next Table.

Weaknesses of the Function Points Measurement Design

Table : Ranges and Ranks for the Internal Logical Files

Attributes		**Rank 1**	**Rank 2**	**Rank 3**
Data element types (DET)	**Intervals**	1 to 19	20 to 50	51+
Record element types (RET)		1	2 to 5	6+

- In Function Points, the intervals within these ranks are not equal, and, of course, the results of a mapping into one of them cannot be added. This classification within one of these ranks leads to an <u>ordinal scale type</u>.

From this point on, it is only possible to say that an observation in rank 2 is greater than an observation in rank 1, but not that it contains twice or three times as many elements.

There is, therefore, a significant loss of flexibility in the mathematical operations allowed.

Weaknesses of the Function Points Measurement Design

- For the sake of clarity, the various ranks (for data elements and record elements) are identified below by the notation used in Table below, that is, ranks D1 to D3 and ranks R1 to R3.

Table 6: Ranks labels

Attribute	**Rank 1:**	**Rank 2:**	**Rank 3:**
Data element types (DET)	D1	D2	D3
Record element types (RET)	R1	R2	R3

Weaknesses of the Function Points Measurement Design

3. The ranks are taken as parameters in a function.
   - This step is more complex from the point of view of a measurement design, as it can be represented in mathematical notation as a function of 2 arguments, , and is illustrated with a 2-dimensional matrix – see Table below.

Table 7: 2-Dimensional Matrix of Ranks

	Data Elements		
Record Elements	Range D1	Range D2	Range D3
Range R1	D1,R1	D2,R1	D3,R1
Range R2	D1,R2	D2,R2	D3,R2
Range R3	D1,R3	D2,R3	D3,R3

Weaknesses of the Function Points Measurement Design

- However, while the parameters in such a function are typically on an absolute or ratio scale type, here the parameters are on an ordinal scale type.
  - Therefore, the positioning of a specific file on the basis of its ranks of DET and RET does not produce a result of an ordinal type

> Even though the contents of cell (D1,R1) can be perceived intuitively as being smaller than the contents of cell (D3,R3), it is not possible to deduce whether the contents of a cell (D1,R3) are equal to, smaller than, or greater than the contents of a cell (D3,R1).

- Therefore, this step corresponds strictly to:
  - the positioning in the matrix on a <u>nominal scale type</u>, with
  - a loss of measurement information, as compared to the ordinal scale of measurement for the individual parameters of this function.

Weaknesses of the Function Points Measurement Design

4. The cells in Table 7 are next segregated into 1 of 3 categories, according to their position in the matrix:
   - Above the inverted diagonal
   - On the inverted diagonal

> The inverted diagonal in Table 7 does not correspond to the usual mathematical notation for a matrix. It does not express a symmetry between the two axes, but rather a simultaneous increase in rank on the intervals.

   - The objective seems to be to express the equivalence of these cells when they are combined on 2 axes. This corresponds to a <u>nominal scale type</u>.

> Example
> The smallest range on one axis combined with the highest range on the other axis yields a category similar to that of the inverse:
> (D1,R3) = (D3, R1)

Weaknesses of the Function Points Measurement Design

5. These 'nominal' categories are assigned the following labels, <u>the semantics of which are of the ordinal scale type</u> – see Table 8:
   - Above the inverted diagonal = **simple**
   - On the inverted diagonal = **medium**
   - Below the inverted diagonal = **complex**

This move from a nominal to an ordinal scale type is not, of course, an admissible mathematical transformation.

Table 8: Assignment of ordered labels

	Data Elements		
Record Elements	Rank D1	Rank D2	Rank D3
Rank R1	Simple	Simple	Medium
Rank R2	Simple	Medium	Complex
Rank R3	Medium	Complex	Complex

Weaknesses of the Function Points Measurement Design

- 6) Weights (or points) are assigned to these ordered labels (simple, medium, complex), as shown in Table 9: these weights for the files are numbers varying from 5 to 15, according to the type of file (Internal or External).

Table 9: Weights for the Data in Function Points

File type	Range 1: Simple	Range 2: Medium	Range 3: Complex
Internal	7	10	15
External	5	7	10

- The intent of this step is clear: to obtain numbers which will be interpreted/perceived as being of a ratio scale type: a complex External File is assigned a weight of 10, that is, twice the weight of a simple one.

This set of weights is the 'trick' used in Function Points to add different functions (Data or Transaction) together (otherwise, the additions would not be allowed, either on the ordinal elements of step 4 or on the nominal elements of step 3).

Weaknesses of the Function Points Measurement Design

7. Finally, the weights (or points) assigned to the files are added together.
   - Under the hypothesis that the numbers obtained previously are on a ratio scale type, the result of the addition will also be a point on a ratio scale type.
   - Of course, such a hypothesis is obviously not supported:
     - This assignation has an incorrect foundation for a ratio scale type. It uses:
       - ranges of irregular intervals for two distinct axes of the matrices,
       - as well as nominal identification within these matrices to which ordinal labels are assigned that are not strictly ordinal values.

Weaknesses of the Function Points Measurement Design

- In conclusion, the mathematical operations used in this set of measurement steps are not all admissible:
  - A number on a nominal scale type (see the steps above) cannot be transformed mathematically into a number on interval or ratio scale type.

- A summary of this analysis of the types of scale in the measurement of the data is presented in Table 10 (Next Slide).

Weaknesses of the Function Points Measurement Design

Table 10: The types of scale in the measurement of the Data in FP

Step	Object	Operation	Scale Type (from)	Scale Type (to)	Mathematical validity	Transformation
1	Data elements	Addition	Absolute	Absolute	Yes	No
	Record elements	Addition	Absolute	Absolute	Yes	No
2	Data elements	Interval identification	Absolute	Ordinal	Yes	Yes and with loss of information
	Record elements	Interval identification	Absolute	Ordinal	Yes	Yes and with loss of information
3	Interval function (data, record)	Position in the matrix	Ordinal	Nominal	Yes	Yes and with loss of information
4	Function of position in the matrix	Name and rank	Nominal	Ordinal	No	Yes and with extra information
5	Function of perceived value	Assign weights	Ordinal	Ratio	No	Yes and with extra information
6	Weights for internal files	Addition	Ratio	Ratio	Yes	No
	Weights for external files	Addition	Ratio	Ratio	Yes	No
7	Internal and external weights	Addition	Ratio	Ratio	Yes	No

Weaknesses of the Function Points Measurement Design

- The measurement process for the logical files calls upon several implicit models, such as those of functional specifications, boundaries between applications, structures, complexity, and weights.
  - None of the relations of these models is based on any experimentally justified theory in any precise framework, either in Albrecht's original paper or in the IFPUG documentation.
    - They are only described in a set of rules established by the normative committee of IFPUG, the Counting Practices Committee.

Weaknesses of the Function Points Measurement Design

- **Analysis of the measurement design for <u>Transactions</u>**
  - All the comments in section 3.2 also apply to the measurement design for the Transactions.
    - Again the transaction measurement processes rely on several implicit models, such as those of documentation, the subdivision of elementary processes and functions, the weights of the transactions, and the selection of weights for the inquiry transactions.
    - Again none of the relations of these models is based on any experimentally justified theory in any identifiable framework, either in Albrecht's original paper or in the IFPUG documentation,.

Weaknesses of the Function Points Measurement Design

- **Addition of all the function types**
  - In the final step of calculating the total number of UFP, all the points for the 5 function types (i.e. the 5 different types of entities: internal logical files, external interface files, input, outputs, and inquiries) are added together.
    - To add such number requires that they be of a ratio scale type.
      - This would be possible if the assignment of weights to the various types of entities had transformed these five different types of entities into a single type of entities, different from the five original components of these entities.
      - Unfortunately, none of this is explained in the Function Points measurement method.
  - While the end results (i.e. the Function Points totals) are considered by the users of Function Points to be numbers on a ratio scale type, they are not derived from the set of mathematical operations embedded within the measurement design of this measurement method.

Weaknesses of the Function Points Measurement Design

Summary of the Analysis

The final results, UFP, are extremely difficult to interpret:

There are so many implicit dimensions and there is so much use of different types of measurement scale types that what seemed simple and intuitively reasonable is a pot-pourri of measurement scale types and of entities which cannot have any mathematical significance, unless the validity of each of the transformations is demonstrated either theoretically or empirically (and this has obviously not yet been done).

Even though the users of Function Points take for granted that FP are on a ratio scale, the nature of these points is unknown:

- What is a Function Point (i.e. What is the measurement unit)?
- What is the intrinsic scale type of Function Points?

Agenda

This chapter covers:

- The origin of Functional Size Measurement
- The global design of Function Points - FP
- Analysis of the design of Unadjusted Function Points
- **Analysis of the design of the Value Adjustment Factor**
- What is a Function Point?
- What if the Function Points weights were dropped?

Weaknesses of the Function Points Measurement Design

- **Analysis of the measurement design for the Value Adjustment Factor (VAF)**
  - The VAF transforms the Unadjusted FP (UFP) into Adjusted FP (AFP), using the 14 General System Characteristics (GSC) of the software application and a linear transformation – Figure 4.
  - Within the design of the VAF: serious methodological weaknesses in the use of quantities with distinct scale types:
    A. The IFPUG Counting Practices Manual provides rules for classifying the software being measured into one of the 5 classifications for each of the 14 GSC.

> When a GSC is not present at all: the characteristic of this software is assigned a classification of 0
>
> When a GSC is at its maximum, the characteristic is assigned a classification of 5.
>
> Various other criteria are provided to assign the intermediate classifications of 1, 2, 3, and 4.

Weaknesses of the Function Points Measurement Design

- This classification from 0 to 5 represents an <u>ordered set</u>, where each value from 0 to 5 is considered greater than the previous one for that characteristic;
  - however, the intervals for the classification are typically:
    - irregular within each characteristic, and
    - different across all 14 characteristics.
- Therefore, the values of 0, 1, 2, 3, 4, and 5 for a characteristic do not represent numbers on a ratio scale, but rather <u>ordered labels, that is, labels with an ordering scale type</u>.

Weaknesses of the Function Points Measurement Design

B. Next, the classification in the previous step of a GSC is multiplied by a degree of influence of 0.1.

The design of this measurement step <u>contains a number of incorrect and inadmissible operations</u>:

a. A multiplication typically requires that the numbers being multiplied be at least on a ratio scale type.
   - This is obviously not the case here: <u>the values of 0 to 5 of the previous steps are not on a ratio scale, but are merely ordered degrees of influence which have no precise quantitative meaning to allow them to be either added or multiplied</u>.

b. Furthermore, the same degree of influence value of 0.1 is assigned to each of the irregular ordered intervals within a GSC.
   - <u>Such an equality (i.e. impact = 0.1 for each interval) across irregular intervals has no theoretically verified justification and is not based on empirical evidence.</u>

c. Finally, while all 14 GSC obviously have different definitions and distinct interval ranges, it could be reasonably argued that they could each have distinct *sets of* degrees of influence, rather than exactly the same set.

Weaknesses of the Function Points Measurement Design

C. In the final step, all the numbers obtained in the previous step for each of the 14 GSC are added together and included in a linear transformation of the VAF to allow an impact of +/-35% to the unadjusted functional size.

- Here again, it can be noted that the measurement design for the VAF calls on several implicit models, such as those of:
    - the set of 14 GSC selected;
    - the 5 interval structures of each of the 14 GSC;
    - the criteria for each interval for each GSC;
    - the equivalence tables for the degrees of influence (that is, 0.1 for each interval);
    - the addition of the various degrees of influence.

- Again, and as mentioned for the data measurement design, <u>not one of the measurement design of each GSC is based on a theory which has been experimentally verified in a well-defined context</u>

Weaknesses of the Function Points Measurement Design

- **Summary of weaknesses in the measurement designs of FP**
  - The various measurement designs within Function Points have been analyzed, and it has been shown that:
    - Several different scale types are used in the various steps.
    - On a strictly mathematical basis, <u>the results of many of the steps and sub-steps of the measurement designs are based on inappropriate use of mathematical properties of corresponding scale types.</u>
  - The consequences are as follows:
    - There is a loss of information and of mathematical flexibility when moving to a lower type of measurement scale (in terms of mathematical properties for the scale type).
    - From a strictly mathematical standpoint, there is an inappropriate use of mathematical properties in the results of many of these measurement steps.
    - There are a large number of unsupported semantical transformations to higher scale types.

Weaknesses of the Function Points Measurement Design

- This, of course, leads to serious challenges in the interpretation of the end results:
  - What is a Function Point exactly?
  - Can Function Points really be treated on a ratio scale, such as is being done by both practitioners and researchers?

Is Function Points a base quantity or a derived quantity?

The structure of Function Points combines a number of steps and a number of base quantities, such as DET, RET, and FTR.

Does this make Function Points a derived measure?

Not really:

– The aim of Function Points is to measure a single concept, that is, functional size.

– The end measurement result in Function Points is not a combination of units, but a single unit, that is, the Function Point, even though such a unit is not formally defined.

Function Points, representing a single concept and having a single unit (i.e. the Function Point), could therefore be considered a base quantity.

Other weaknesses

- **Mix of semantics: functional size, technical size, and complexity**
  - The ISO requires that FSM methods quantify only the functional size of software, without taking into account its technical and quality characteristics [ISO 14143-1].
    - On the one hand, the 14 GSC of Function Points definitively take into account a number of the technical characteristics of software, such as performance, response-time, reuse, etc.
    - Therefore, the ISO does not recognize any of these as meeting the ISO requirements, and only the unadjusted part of Function Points has been adopted as an ISO standard.

Other weaknesses

- On the other hand, within the Function Points design, it is explicitly stated that the data (and transactions) are being classified on a complexity scale embedded within its tables of weights.
  - Since complexity can be considered as part of the quality of software, it can be easily argued that the Function Points design includes a quality characteristic of the software.
    - This was not addressed when Function Points was submitted to the ISO with the claim that it met all the mandatory requirements of ISO 14143-1.

Other weaknesses

- **Implicit relational system**
  - The design of Function Points was based on expert judgments, and it is described in terms of rules, rather than in terms of models and measurement principles.
  - Some of the implicit models of Function Points are listed below (this list is by no means exhaustive):
    - a model of the user's perspective
    - a model of 5 types of functions
    - a model of the logical file type (internal and external)
    - a model of the transaction function type
    - a sub model of transactions (add, modify, delete)
    - a model of elementary components
    - a model of the structure of the decision tables
    - a model of the weights in the above structures

Other weaknesses

- It is therefore necessary to clarify the domain of the relations or the measurement space of Function Points [Fenton 1991], and the domain of the measured relations must be explicitly defined if we want to use it in an appropriate manner and possibly to modify it in order to extend its domain of application.
  - Since there are no explicit models, it is not easy to examine the fundamental principles of the structure of Function Point Analysis (FPA), and its credibility as a *bona fide* measurement method has suffered as a result.

- This absence of explicit models has made it next to impossible, in practice, to foster an evolution over the past 30 years of the numerical assignment rules and of the structure of the official version of Function Points.

Agenda

This chapter covers:

- The origin of Functional Size Measurement
- The global design of Function Points - FP
- Analysis of the design of Unadjusted Function Points
- Analysis of the design of the Value Adjustment Factor
- **What is a Function Point?**
- What if the Function Points weights were dropped?

What is a Function Point?

- **The original definition of Function Points**
  - Albrecht's initial intention was to measure productivity, and to do this he had to define and measure an output (i.e. the software product developed) and an input (i.e. the effort).
    - He defined this software output as "*function value delivered*" and his objective was "*to develop a relative measure of function value delivered to the user that was independent of the particular technology or approach used*" [Albrecht, 1979].
  - The result of that methodology in terms of his objective had given him, in his words, a "*dimensionless number defined in function point(s)*" [Albrecht, 1979].

What is a Function Point?

- There is, however, a certain contradiction in expressing the size of an application in terms of a dimensionless number, since every measurement system necessarily depends on a reference system which allows us to interpret the measurements [Fenton 1991].

> It must be remembered that a measurement is not a number in itself, but the assignment of a number: it is the establishment of a relation between the entities and the characteristic under observation.
>
> Every time a characteristic is measured, it must be in terms of a specific set of relations.

  - It seems that, because Albrecht stated that Function Points did not have any dimension, this assertion has not been challenged, either by researchers or by practitioners.
  - In this context, there is room to revisit the significance of FP as a measurement concept and to examine what Function Points mean from the point of view of measurement.

What is a Function Point?

- ***Revisiting the original definition***
  - The existence and true description of the system of relations cannot be deduced just from Albrecht's incomplete definition;
    - it must be deduced from the reference context and the selection criteria of the experiment he used to design FP.
  - Therefore, it becomes essential to clarify the interpretation of FP in relation to traditional metrology and in relation to its initial application in Albrecht's empirical model.

What is a Function Point?

- The analysis of Albrecht's intentions (i.e. his implicit measurement model) and of the reference system he used to specify the structure and the parameters of FP must therefore serve as a basis for an interpretation of his initial definition, from the perspective of a measurement system in the metrology sense.

- It is important to properly understand the reference context in which a function type is developed. If the composition of the reference context is known, it is then possible:
  - to analyze it, and
  - to define new contexts in terms of the reference context.

- This is what allows a reference function type to be moved from one context to another.

What is a Function Point?

- ***Function Points relational system***
  - To identify and analyze the domain of the relations in FP, it is necessary to go back to its original sources.

> The empirical model of FP, including the procedures and weights, can be traced back to Albrecht's development environment and to his criteria for selecting the projects to be measured, as described in his initial 1979 paper.

  - Albrecht's 1979 paper describes:
    - his global context of measurement (Table 2), as well as
    - the criteria for selecting projects to be measured – Table 3.
  - This Albrecht's paper therefore documented, and defined, the reference context in which, and for which, FP had been built.

What is a Function Point?

Table 2: Albrecht 1979 – Description of the Initial Reference Context

- The organization includes 450 people who develop applications.
- Development is under contract to clients of IBM.
- The developers and the clients are dispersed throughout the US.
- At any time, between 150 and 170 contracts are open.
- An average contract involves two or three people.
- Each contract is made in the context of a specific development framework.
- The majority of the contracts are limited to certain phases of the methodology.
- Based on their experience, the design phase accounts for 20% of the work hours, the implementation for 80%.
- It is necessary to measure the whole process, including the design and the costs incurred during the design.
- The projects were completed between mid-1974 and the start of 1979.
- The size of the projects varies from 500 to 105,000 hours of work.
- Of about 1500 contracts for the period, only 22 met the selection criteria.

What is a Function Point?

- The analysis of this initial reference context allows us to clarify the domain of the mapping of the initial empirical model which formed the starting point for the rules and procedures:
  - Albrecht derived his measurement space (i.e. model of relationships) using the criteria enumerated in Tables 2 and 3.
  - This set of criteria defined a stable and relatively homogeneous set of development conditions (i.e. his experimental design), thus explicitly limiting the number of outside influences on the development process as much as on the software products analyzed.

Table 3: Albrecht 1979 – Selection Criteria for Admissible Projects

1	The project must have passed through all phases of the methodology (from definition of requirements to implementation) and must have been delivered to the client.
2	The whole project must have been properly managed with consistent definitions of the tasks and the processes of management.
3	All the hours of work put in by IBM and by the client must be known and must have been carefully measured.
4	The functional factors must be known.

What is a Function Point?

- Now, if we examine this system of relations in which FP was defined, it becomes obvious that the notion of effort (or productivity) has been present from the beginning with the introduction of weights, which were set by ‘debate and trial’.

- Albrecht defined:
  - an initial *reference context* as part of the process that led to his selecting 22 observations to measure from among the 1,500 observations available to him during this period – Table 2;
  - *selection criteria* based on how much information was available to describe the effort required for the complete development cycle for the selected projects – Table 3.

What is a Function Point?

- We may observe, therefore, that his selection of projects depended (for the selection of weights) on a productivity model based on knowledge of
  - a complete and homogeneous development process – Table 2, and
  - all the efforts required to complete a full development cycle – Table 3.

- The quality of this system of relations can be deduced from the analytical methodology used by Albrecht [1979 and 1983] to study the relation between effort and Function Points in his own reference context (i.e. his empirical design).

What is a Function Point?

- Albrecht used the statistical technique of linear regression to build his productivity model, quantifying the relationship between the dependent variable (effort) and the independent variable (functional size in Function Points).
  - He obtained a linear model with a fairly high coefficient of determination of R^2 of 0.87 (the maximum of an R^2 being 1.0).

Influence of outliers in regression models

[Knaff 1986] and [MacDonell 1991] have pointed out that Albrecht's dataset contained three projects which were much larger than the others, and that these three projects served as anchor points in the statistical regression (these three points should then be considered as outliers and excluded from further analysis).

When these three outlier projects are excluded from the building of the regression model, the value of the coefficient of determination (R^2) falls to 0.42 for the other nineteen projects in Albrecht's sample.

What is a Function Point?

Table 12: Empirical relation between FP and Effort in Albrecht's dataset

Dependent Variable	**Independent variable**	**Comment**	**R^2**
Effort	Adjusted FP	All 22 projects in the 1979 paper (* 24 projects in the 1983 paper)	0.869
Effort	Adjusted FP	Excluding the 3 largest projects [Knaff et al., 1986]	0.42

What is a Function Point?

An interpretation of a Function Point

- To gain insight into the initial FP structure, it is necessary to distinguish between various concepts (such as the measurement of size, the measurement of effort), and the measurement of the general relation between them, <u>expressed in the form of implicit or explicit productivity models</u>.

- In terms of the analysis of the initial reference context described in the previous section and the implicit system of relations, if the context is to be changed, the transformation must be a linear one.
  - It is therefore desirable to revise the initial definition of a Function Point and make it more precise, using the following interpretation:

Interpretation of a Function Point

Function Points constitute a system of relations between:
— the measurement of a "reference function type", and
— the measurement of the effort in the initial reference context established by Albrecht.

ISO 14143-1 requires that functional size in an ISO-recognized measurement method not be related to effort.

The above interpretation of a Function Point challenges the assumption that FP meets this mandatory requirement of ISO 14143-1.

What is a Function Point?

- Unfortunately, the term ‘reference function type’ is not explicit in Albrecht’s work:
  1. It is implicit only, and it has to be inferred from:
     - the rules for identifying functions, and
     - the rules for identifying and counting the elementary components of these functions.
  2. The implicit system of relations between the measured value of a reference function type and the measured value of the effort is based on weights, as well as on Albrecht's algorithm for assigning them
     - i.e. the system of implicit relations that would justify – <u>should such justification be properly documented</u> – the scale-type transformations included in the algorithm and the weights.

Agenda

This chapter covers:

- The origin of Functional Size Measurement
- The global design of Function Points - FP
- Analysis of the design of Unadjusted Function Points
- Analysis of the design of the Value Adjustment Factor
- What is a Function Point?
- **What if the Function Points weights were dropped?**

Other Research Findings

- **What if the FP weights were dropped?**
  - What would happen, if the pot-pourri of scale types with inadmissible mathematical transformations and weights were dropped from this measurement design?
    - Would it still be possible to obtain a reasonably good productivity (and estimation) model using only the base components of the Function Points measurement method?

Other Research Findings

- This specific issue was explored with an industrial dataset in Abran & Robillard [1996].

- Various subsidiary questions were also addressed, for example:
  - Is the conventional FP model (i.e. its 5 types of functions, all combined) better than the sub-models, where the functions are taken individually or combined in a different way?
  - How relevant and useful are the weights?
  - How relevant and useful are the algorithms?
  - Do any of the steps lower the quality of the relation between size and effort?

Other Research Findings

- The key findings from the empirical research with this industrial dataset can be summarized as follows:
  - Productivity models built using only the base components (i.e. DET, RET, FTR, and various combinations of these components) are as good as models built with the full FP design (with weights and algorithms).

- The corollary of this key finding is:
  - The weights and algorithms do not contribute much to the quality of the productivity models built with this dataset.

Other Research Findings

- Said differently, these weights and algorithms could be considered as <u>'feel good' artifices</u>:
  - The fact that they are included within a measurement method leads the practitioners to *believe* that many details have been duly and adequately taken into account in the analysis of the effort relationship.
    - However, since the mathematical structure that handles these artifices is improper, none of these weights and algorithms brings a significant contribution to the effort relationship.
    - Therefore, <u>these artifices give practitioners a false sense of security, even though they 'feel good' about them</u>.

These findings on weights are also important to researchers

Attempting to improve the <u>set of weights or the algorithms</u> does not address the fundamental issues of the design of measurement methods based on a design including such exogenous factors and weights.

Other Research Findings

- **What if the Value Adjustment Factor is dropped?**

  - We have seen, in section 3.5, the numerous mathematical issues hampering the correct treatment of the 14 GSC of Function Points.
    - What happens when these 14 GSC of Function Points are dropped in the construction of productivity models?

  - A number of researchers have investigated this issue by building estimation models with both AFP and UFP.
    - Findings from these empirical studies lead to the conclusion that models built <u>without</u> the 14 GSC of Function Points <u>are as good as those taking them into account.</u>

Other Research Findings

- Said differently, the VAF built from the 14 GSC are 'feel good' artifices that lead practitioners to *believe* that many cost factors have been taken into account in the analysis of effort relationships:
  - Since the mathematical structure that handles these artifices is improper, none of these brings a significant contribution to those relationships.
- In summary, the GSC artifices give practitioners a false sense of security that so many factors have indeed been taken into account.

These findings on GSC are also important to researchers

Attempting to improve the set of 14 GSC of Function Points is not addressing the fundamental issues associated with measurement methods in which such exogenous factors are used in their design.

Agenda

This chapter has covered:

- The origin of Functional Size Measurement
- The global design of Function Points - FP
- Analysis of the design of Unadjusted Function Points
- Analysis of the design of the Value Adjustment Factor
- The interpretation of what is a Function Point on the basis of the experimental context of its design in the late 1970s
- Analysis of the impact of not taking the weights into account in the Function Points design