Function Points Analysis: An Empirical Study of Its Measurement Processes

Alain Abran, *Member, IEEE*, and Pierre N. Robillard, *Member, IEEE*

Abstract—Function point analysis (FPA) was initially designed on the basis of expert judgments, without explicit reference to any theoretical foundation. From the point of view of the measurement scales used in its measurement process, FPA constitutes a pot-pourri of scales not admissible without the transformations imbedded in the implicit models of expert judgments. The results of this empirical study demonstrate that in a homogeneous environment not burdened with major differences in productivity factors there is a clear relationship between FPA' primary components and Work-Effort. This empirical study also indicates that there is such a relationship for each step of the FPA measurement process prior to the mixing of scales and the assignments of weights. Comparisons with FPA productivity models based on weights confirm, on the one hand, that the weights do not add information and, on the other, that the weights are fairly robust and can be used when little historical data is available. The full data set is provided for future studies.

Index Terms—Function point analysis, productivity models, estimation models, measurement process, functional metrics.

1 Introduction

FUNCTION point analysis (FPA), developed by Allan Albrecht of IBM, was first published in 1979 [1], and, in 1984, the International Function Point Users Group (IFPUG) was set up to clarify the rules, set standards, and promote their use and evolution. FPA provides a standardized methodology for measuring the various functions of a software application. FPA measures functionality from the user's point of view, that is, on the basis of what the user requests and receives in return.

The function points (FPs) are obtained by measuring the software application from two distinct perspectives (Fig. 1):

1) The functional size, calculated by assigning weights to *each individual function*. This will be referred to as the functional size measurement process which produces the unadjusted function points (UFP). This perspective includes both a data function type measurement process and a transaction function type measurement process.
2) The value adjustment factor (VAF), calculated using predefined general systems characteristics (GSC) to assess the environment and processing complexity of the software application *as a whole*. This will be referred to as the adjustment measurement process.

The value adjustment factor in 2) adjusts the functional size determined in 1) to produce the adjusted function points (AFP). This is referred to in Fig. 1 as the function points measurement model.

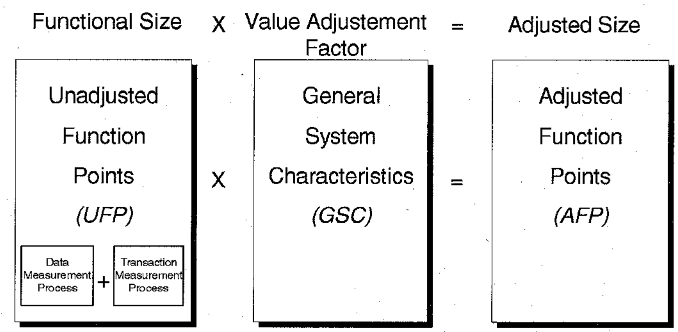


Fig. 1. Function points measurement model.

Over the years, various refinements have been made to the 1979 initial description and, up until 1996, seven successive versions had been published (Table 1). The first three addressed the structure of FPA, while the four IFPUG[1] versions provided clarification of the rules and counting guidelines.

TABLE 1
SEQUENCE OF OFFICIAL FPA VERSIONS

1	Albrecht 79 [1]
2	Albrecht 83 [2]
3	GUIDE 84 [3]
4	IFPUG 86 [4]
5	IFPUG 88 [5]
6	IFPUG 90 [6]
7	IFPUG 94 [7]

The 1979 model had four function types, one set of function weights (Table 2, left-hand side) and 10 application general systems characteristics (GSC), for a maximum value adjustment factor of ±25%. The 1983 model was expanded to five function types, three sets of weights for each function type (Table 2, right-hand side) and 14 application general systems characteristics, for a maximum value adjustment factor of ±35%.

1. IFPUG: International Function Point Users Group.

• A. Abran is with the Université du Québec à Montréal, P.O. Box 8888, Succ. Centre-Ville Montreal, Quebec, Canada H3C 3P8. E-mail: abran.alain@uqam.ca.
• P.N. Robillard is with École Polytechnique de Montréal, P.O. Box 6079, Succ. Centre-Ville Montreal, Quebec, Canada H3C 3A7. E-mail: robillard@rgl.polymtl.ca.

Manuscript received July 12, 1992; revised Oct. 11, 1996.
Recommended for acceptance by B. Littlewood.
For information on obtaining reprints of this article, please send e-mail to: transse@computer.org, and reference IEEECS Log Number S95090.

0098-5589/96$05.00

TABLE 2
ALBRECHT 79 AND 83-WEIGHTS

Albrecht 79			Albrecht 83				
Function Types		Weights	Function Types		Low	Average	High
1	Files	10	1	Internal logical files	7	10	15
			2	External interface files	5	7	10
2	Inputs	4	3	External inputs	3	4	6
3	Outputs	5	4	External outputs	4	5	7
4	Inquiries	4	5	External inquiries	3	4	6

TABLE 3
EXAMPLE OF A FUNCTION POINT COUNT - ALBRECHT 83 VERSION

Function Types	No. Functions * Weights = UFP	Complexity Adjustment Factor
Internal logical files	3 * 10 = 30	GSC 1 to 11 = .00
External interface files	0 * 7 = 0	
External Inputs	2 * 4 = 8	GSC 12 to 14 = .04 each
External Outputs	2 * 5 = 10	
External Inquiries	5 * 4 = 20	**Total** = .12
Total	**UFP = 68**	VAF = (.65 + .12) = .77
Adjusted Function points = UFP * VAF = 68 UFP * (0.77) = **52 AFP**		

TABLE 4
FPA OF PRIMARY COMPONENTS--IFPUG 1994

Types	Abbreviation	Definition
Data Element Type	DET	A unique user recognizable, nonrecursive field on an internal logical file or external interface file.
Record Element Type	RET	A unique user recognizable subgroup of data elements within an internal logical file or external interface file.
File Type Referenced	FTR	An internal logical file or external interface file read or maintained by a transactional function type.

An example of a project count using the Albrecht 83 version is presented in Table 3. The software measured in this example contains three internal files, two inputs, two outputs, and five inquiries, for a total of 68 unadjusted function points (UFP) when all functions have an average weight classification. The total of the general system characteristics is calculated as 0.12 (Table 3, right-hand side), with GSC 1 to 11 having no influence (value = 0) and GSC 12 to 14 rated as significant with a degree of influence of 0.04 each. To get the value adjustment factor (VAF), this adjustment factor of 0.12 is then added to 0.65 to fall within the adjustment range of ±35%. These two results, UFP and VAF, are then combined to produce an adjusted size of 52 adjusted function points (AFP) for this simplified example (Table 3, bottom line).

In the Albrecht 83 version, however, the assessment of function-type complexity (low, average, or high) was a subjective process. To transform this assessment into an objective process consistent across individuals and organizations, the GUIDE 84 version introduced a new dimension to FPA: the function types were decomposed into primary components and two-dimensional matrices with predetermined ranges of values were used for classification purposes. Table 4 presents the definitions of FPA primary components, and Table 5 presents the two-dimensional matrix designed for the data-type functions (internal logical files and external interface files).

The four subsequent versions published by IFPUG provided further clarification of the rules, guidelines and criteria,

TABLE 5
MATRIX STRUCTURE DESIGNED IN GUIDE 84 FOR DATA-TYPE FUNCTIONS

Record Element Types (RET)	Data Element Types (DET)		
	1 to 19	20 to 50	51+
1	L	L	A
2 to 5	L	A	H
6+	A	H	H
L = Low; A = Average; H = High			

but did not introduce any change to the structure itself of FPA measurement process. The current official version, IFPUG 94, still uses the Albrecht 83 function types and weights, as well as the GUIDE 84 matrices.

Most publications on FPA have addressed issues that are not related to its structure, such as comparisons with other software metrics based on lines of code [2], the accuracy of estimates [8], the consistency of the counts when made by different counters or the inter-rater reliability [9], [10], [11], [12], and productivity analysis [8], [13], [14], [15], [16]. Only a few authors have reviewed FPA methodology and identified some of its weaknesses in areas such as domain of applicability, the structure of its primary components, and the impact of the adjustment algorithm [17], [18], [19], [20].

Function points were described by Albrecht as follows: "a *dimensionless number defined in function points, which we have found to be an effective relative measure of function value delivered to our customer*" [1]. However, there is a descriptive dissonance in saying that the *size* of an application can be expressed through a *"dimensionless number."*

Measures should be supported by the discipline of measurement theory and a measure should always represent a mapping to a specific model: An attribute can be measured if there is a mapping from an empirical relation system into a numerical relation system [21].

An implicit hypothesis in Albrecht's original definition of function points and in his empirical study with 22 projects seems to be that it explains the relationship of the work effort required to deliver these functions. In his empirical study, the work effort is used as the independent variable in lieu of the *"value delivered to the customer."* This has been made possible through the assignment of weights to the functions and certainly qualifies as an implicit transformation in the size measurement of a software application.

The research issues discussed in this paper can then be stated as follows: Where does FPA stand with respect to measurement systems, and what are the validity and impact of each of its measurement step with respect to the work-effort relationship? The domain of this analysis is limited to these issues, within the constraints of the empirical designs, as opposed to the more global issue of general productivity and estimation models.

In this paper, function point analysis (FPA) and function points (FP) interpretation are reconsidered from a measurement perspective and the issue of the implicit models is addressed, as well as the measurement process itself with respect to the measurement scales and transformations in all the measurement steps. In this research work, there is no attempt to describe these relationships. The research objective is strictly limited to verifying the external manifestation of the existence and impact of the implicit transformations.

Section 2 presents an analysis of FPA from a measurement perspective. It includes an analysis of the measurement steps in the *data* measurement process: for each measurement step, the scale is identified and verified against the admissible mathematical transformations.

Section 3 presents a summary of published empirical evidence of the relationship between function points and work effort. Section 4 presents the empirical designs: the historical database and the methodology for the study of projects that qualify for a homogenous development environment unburdened by major differences in productivity factors. Section 5 present the statistical analysis of productivity models based on the primary components of FPA, prior to the mixing of scales. This is followed in Section 6 by analyzes of productivity models built with variables which takes into account the mixing of scales through the assignments of weights.

The interpretation of the results is presented in Section 7 which suggests that each FPA step adds information and represents transformations which maintain, and often improve, the relationship with respect to work effort.

2 Analysis of FPA Measurement Process

The objectives in this section are twofold: to identify which measurement scales (nominal, ordinal, interval, and ratio) are integrated in the formal measurement process of FPA, and to analyze how they are transformed through the various steps of this measurement methodology. This approach was also applied in [22],[23] for the analysis of lines of code complexity metrics. This section presents a summary of results discussed in [24], [25], [26].

FPA is modeled from a measurement perspective in Fig. 1: the *data* measurement process, the *transaction* measurement process, and the *adjustment measurement process.* These are themselves complex processes [25]. Only the data measurement process is used for illustrative purposes in this section.

2.1 FPA Data Measurement Process Model

The data measurement process is decomposed into five steps (Fig. 2):

F1). The first step takes as input the application / project documentation and produces as output the *list* of *logical files.* It should be noted that while this measurement view is described in the FPA measurement methodology in terms of rules and guidelines, its basic relationships have not been formally analyzed.

F2). The second step analyzes the boundary between the software application / project being measured and either external applications or the user domain in order to classify the logical files into internal logical files or external interface files. The output of this process consists of sublists of *infernal logical files* (ILF) and *external interface files* (EIF).

F3). The third step counts the actual number of *data element types* (DET) and *record element types* (RET) within each file type, as defined in Table 4. Here, for the sake of clarity, the term "data" refers to DET, and "record" refers to RET.

F4). The fourth step applies the *data algorithm* with the following inputs: data counts (DET), record counts (RET), data matrix structure (Table 5) and data weights (Table 2). The output of this process is a *list of points* for all logical files.

F5). The last data measurement step consists in adding all points from step F4) to produce the *unadjusted data count.*

2.2 Measurement Scales Analysis

The measurement scales and the mathematical operations permitted for each scale are summarized as follows.

1) *Nominal Scale.* This scale is used to name objects or events for identification purposes only, and there are no quantitative implications associated with it. Only nonparametric statistics can be used.
2) *Ordinal Scale.* This scale is used to order or rank items, based on a criterion that can be subjective or, preferably, objective. Rank order statistics and all that apply to the nominal scale can be used.
3) Interval Scale. This scale is used to determine the difference between the ranks; it is continuous between two end-points, neither of which is necessarily fixed. With this scale, the items can be distinguished and ranked, and the differences between ranks measured. Arithmetic mean and all statistics that apply to the ordinal scale can be applied.

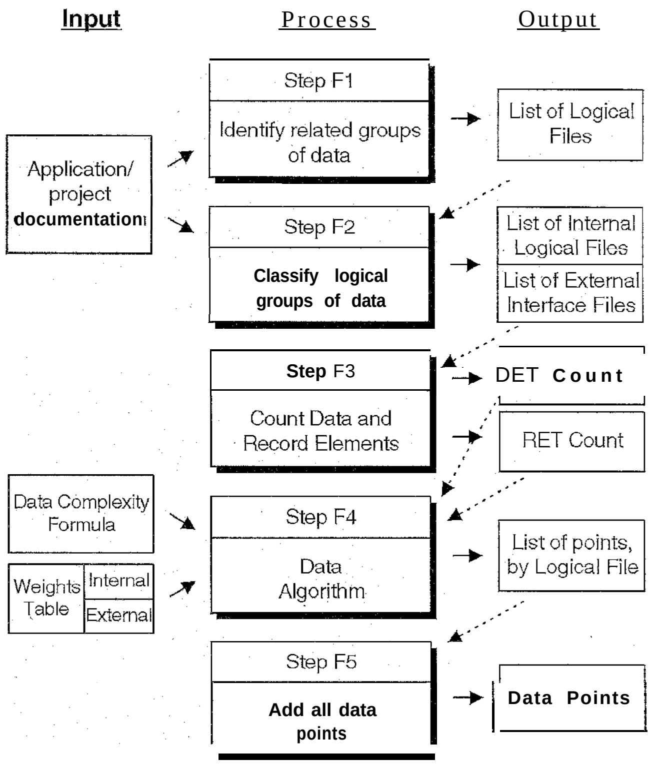


Fig. 2. Data measurement process

Ratio Scale. With this scale, no negative values can be used to multiply measurement values. Percentage calculation and all statistics that apply to the interval scale can be applied.

Absolute Scale. In addition to the properties of the ratio scale, the absolute scale has a unique origin from which to begin the measurement. Within this scale, entities can be counted.

The uses of the measurement scales are examined by means of the data measurement process shown in Fig. 2. In FPA methodology, the algorithm of Step F4) for the 'calculation of points is usually described as a two-step calculation: a file is first classified as being of low, average, or *high* complexity, then a number of points (weight) is assigned depending on this level of complexity.

For example, a software application has three internal logical files as the output of step F2) and the following characteristics as the output of step F3):

Internal logical file 1:	has one record (RET) and four data elements (DET),
Internal logical file 2:	has two records (RET) and 21 data elements (DET),
Internal logical file 3:	has six records (RET) and 26 data elements (DET),

In Step F4), file 1 is classified as low based on Table 5 of the FPA methodology and, based on Table 2 (Albrecht 83-weights), a weight of 7 is then assigned; similarly, file 2 is classified as average with a weight of 10 and file 3 is classified as *high* with a weight of 15. The addition of these weights gives a total number of 32 points for the three internal logical files described above.

This descriptive procedure is, however, an oversimplification of the FPA measurement process. In order to identify the types of scales and to analyze their uses in the data measurement process, step F4) has been further broken down into four substeps in [25] for the purpose of identifying the measurement transformations occurring in the utilisation of the data matrix structure and the assignment of weights.

The results of the scale analysis are summarized in Table 6: For each step, the measurement objects and the mathematical operations on these objects are identified (columns 1 to 3). The next two columns indicate the type of scale of the objects prior to the mathematical operation (column 4—*Scale:* From) and after (column 5-Scale: To). The second column from the right indicates the mathematical validity of the operation: it is valid to stay within the same scale or to move to a scale with fewer mathematical properties, for example, from an ordinal scale to a nominal scale, as in step F4b). The right-hand column indicates if a change of scale has occurred within the step and, if so, whether some information is being lost or added within the transformation that has occurred. This last column indicates only that such an implicit transformation has occurred, but does not describe it.

It can be observed that the measurement scales from step F3) to step F4b) move from the absolute scale to the ordinal scale and then to the nominal scale. These transformations are mathematically admissible, but with a loss of measurement information and mathematical flexibility: while the information in step F3) can be added, the only mathematical operation still admissible in step F4b) is the ability to identify and name (without ranking or addition properties).

However, step F4c) moves in the opposite direction along the measurement scales. The transformation from a nominal scale to an ordinal scale is not derived from the mathematical properties only.

Similarly, step F4d) moves in the opposite direction along the measurement scales: it assigns a weight to a rank, based on whether it is an internal logical file or an external interface file. These weights represent relative absolute numbers from 5 to 15, depending on the data function type, and the end number obtained in this step is taken as a ratio number in subsequent steps. Again, this transformation from an ordinal scale to a ratio scale is not derived from the mathematical properties only.

Finally, in steps F5a) and F5b), the results of the assignments of weights to each data function are added together.

All the above comments on the uses of the scales apply to the *transaction* measurement process and to the adjustment *measurement process* of Fig. 1. In the final step of FPA measurement methodology, all the points are added together, whether they come from internal logical files, external interface files, inputs, outputs, or inquiries. It must be noted that this has been made possible only by the assign-, ment of the weights to transform five different types of objects into one single object of a different type and of an unspecified nature, that is, a function point> (FP). Again, this is done only through implicit transformations, and not through a strict measurement process with the proper use of the different types of scales.

TABLE 6
USES OF SCALES IN THE DATA MEASUREMENT PROCESS

Step	Objects	Operation	Scale: From	Scale: To	Math. validity	Implicit transformation
F3)	Data	Count	Absolute	Absolute	Yes	No
	Record	Count	Absolute	Absolute	Yes	No
F4a)	Data	Identify range	Absolute	Ordinal	Yes	Yes and loss of information
	Record	Identify range	Absolute	Ordinal	Yes	Yes and loss of information
F4b)	Function of ranges of (data, record)	Position in matrix	Ordinal	Nominal	Yes	Yes and loss of information
F4c)	Function of position in matrix	Name and order	Nominal	Ordinal	No	Yes and addition of information
F4d)	Function of perceived values	Assign weights	Ordinal	Ratio	No	Yes and addition of information
F5a)	Weights of internal files	Add	Ratio	Ratio	Yes	No
	Weights of external files-	Add	Ratio	Ratio	Yes	No
F5b)	Weights: internal + external	Add	Ratio	Ratio	Yes	No

The end results (unadjusted and adjusted function points) become, therefore, very difficult to interpret: there are so many dimensions involved and so many uses of different types of scales that the end measure, which might look rather simple and reasonable, is, in fact, a pot pourri that might not have correct mathematical meaning. This confirms Eijogu's assertion that the end results may not be mathematically admissible, especially with respect to units and dimensions [27].

However, the above analysis of the measurement scales of FPA has identified the existence of implicit transformations (relationships) without which the FPA measurement results would be invalid. The FPA measurement methodology does not derive from a well-defined and proven theory: currently, it is entirely empirically based on expert opinion.

The mapping, or measurement space, of FPA undoubtedly needs to be clarified if it is to be trusted as a valid measurement system. The domain of relationships being measured must be made more explicit if it is to be used properly, and possibly modified to expand its domain of applicability. In the next sections an approach is proposed to explore FPA measurement space and the domain of the relationships being measured.

3 EMPIRICAL EVIDENCE OF THE FP/WORK-EFFORT RELATIONSHIP

Albrecht's definition quoted in the introduction states that Function points are *"... found to be an effective relative measure of function value delivered to our customer."* This was demonstrated in his research papers [1], [2] through an analysis of the relationship between function points and work effort.

Various other researchers have verified that there is indeed a strong empirical relationship between the size of an application measured with function points and work effort [2], [8], [14], [15], [17], [18], [28]. Table 7 presents a summary of the regression coefficients (R^2) of work effort relative to function points. Kemerer's and Desharnais' results concur in that they fall within the 0.50 to 0.55 range for the R^2, while Albrecht's results are much higher at 0.869. It was pointed out [29], however, that three projects in Albrecht's sample were particularly large compared to the others, and that without these three projects, the R^2 would be lower by 0.44 to 0.42. It should be noted, too, that the value adjustment measurement process using the 14 general system characteristics does not add much to the R^2 in either the Kemerer or the Desharnais sample.

Empirical study on FPA and its relationship to work effort has focused exclusively on the end product of the measurement process: the total count of either the unadjusted or adjusted function points. Little empirical research has included data on the intermediate steps of the FPA measurement methodology. From the analysis of the measurement steps and the empirical results on the FP/work-effort relationship, an additional research question is defined as follows:

> If FPA is a measurement system, rather than simply a recipe for' constructing a dimensionless number, then each of the measurement steps, from the beginning to the end result, has a specific meaning and contributes to the measurement process.

If this is so, each of the measurement steps adds information. Furthermore, if each step is valid, then not only is the end result of the last step useful, but the intermediate steps could also hold similar relationships. Some of the intermediate steps might possibly be even more meaningful if, in subsequent steps, information is lost and not added.

4 EMPIRICAL DESIGNS

4.1 Sets of Independent Variables

In order to analyze both the validity and the impact of the transformations in each of the measurement steps, 10 sets of independent variables have been identified for the design of productivity models based either on FPA primary components or on the FP counts after the mixing of the scales. These sets are listed in Table 8, together with the abbreviation of each variable.

TABLE7

EMPIRICAL RESEARCH ON THE FP/WORK-EFFORT RELATIONSHIP

Models	Independent variable	Comments	R^2
Albrecht 83	Adjusted FPs	All 22 projects	.869
Albrecht 83	Adjusted FPs	Excluding three largest projects [29]	.42
Kemerer 87	Adjusted FPs Unadjusted FPs		.553 .538
Desharnais 88	Adjusted FPs Unadjusted FPs		.540 .498
Emrick 88	Adjusted FPs	*Correlation Factor, instead of R^2	.66*

TABLE 8

SETS OF INDEPENDENT VARIABLES FOR THE EMPIRICAL DESIGNS

Basis	**No.** of independent variables	The independent variables and their abbreviations
Primary components	1 variable	**Total number of data** elements (TDET)
		Total number of logical groups of data elements' (TGRE)
	2 variables	Total number of data elements and groups of elements (TDET + TGRE)
		Subtotal of data and groups of data, in the Data Measurement Process only (SDET-D and SGRE-D)
		Subtotal of data and groups of data, in the Transactions Measurement Process only (SDET-TR and SGRE-TR)
	4 variables	Subtotals of data and groups of data, in both the Data and Transactions Measurement Processes
	10 variables	Subtotal of data and groups of data within each of the five function types (DETIF-data element for internal files, GREIF-groups of elements for internal files, ...)
Mixing of scales: weights	1 variable	Adjusted Function points (AFP)
		Unadjusted Function points (UFP)
	5 variables	Subtotal of Function points for each of the five function point types (FP-IF: points for internal files; FP-EF: Function points for external files; FP-IP: input; FP-OP: output; and FP-IQ: inquiry)

**Logical groups of elements = FTR for the transaction-type functions or RET for the data-type functions (GRE).*

Using the primary components as a basis, five sets of independent variables were identified. The first two have only one independent variable: Either the total number of data elements (the DET on the horizontal axis of the FPA matrix structure of Table 5), or the total number of logical groups of data elements (the RET on the vertical axis of the FPA matrix structure of Table 5). The third set includes both data elements and logical groups of data elements into a 2-variable set. The next three sets look into the two different function-type groups-data and transactions, first individually with Z-variable sets, and then combined into a 4-variable set. The last set using FPA primary components as its basis is a lo-variable set which investigates the implicit model of five function types (internal logical files, external interface files, inputs, outputs, and inquiries)

The next three sets of independent variables (lower portion of Table 8) includes variables based on the FPA formula after its mixing of scales and assignment of weights. The first two sets are the standard l-variable models found in the literature with either the unadjusted Function points (UFP) or the adjusted Function points (AFP) as the independent variable. The last set is a 5-variable model which utilizes the subtotals of UFP for each of the five function types.

For our empirical designs, each set will then be used to derive a productivity model based on the least-square regression technique using the SAS statistical package. For each of the regression models derived, the following four statistics will be presented: the standard error of the estimate (expressed in number of days), the coefficient of variation with respect to the actual mean of the dependent variable (expressed in percentage), the coefficient of single or multiple determination R^2, and the adjusted R^2 (the latter is more relevant when there is a large number of independent variables). The productivity models derived from these sets will then be used to verify the existence of the implicit relationships, and the degree of them, throughout the FPA measurement process.

4.2 Empirical Data Set

An historical database of 37 projects (1986-1990) of a major Canadian financial organization was used to carry out the empirical designs. During the data collection period, this. organization was using the FPA rules as documented in the work-in-progress of the IFPUG standards committee, which were later published in the IFPUG 90 version [6]. All of the IFPUG 90 rules were applied in the data collection process, with the following exception: the primary components were identified based on the entity/relation model described in [7], [17].

This data set is illustrated in Fig. 3, with the unadjusted Function points on the x-axis (min: 39 FP, max: 1,542 FP), and the actual effort-days on the y-axis (min: 52 days, max: 924 days). The average work effort (WE) size of these 37 projects is 362 days and the average functional size is 213 FP.

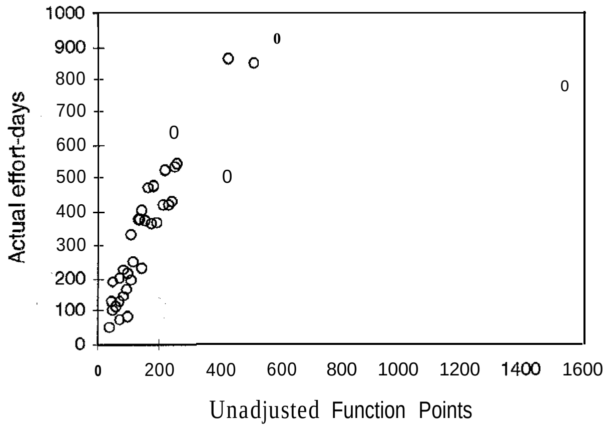


Fig. 3. The full data set of 37 projects.

The model adequacy was analyzed, or the explanatory power of the independent variable in accounting for the variability of the dependent variable, typically measured by the coefficient of determination R^2 [35]. The coefficient of determination R^2 with the adjusted function points (AFP) as the independent variable is 0.47 for the data set of 37 observations (Table 9). This value of the R^2 can be compared to published results, as illustrated in Table 7. The standard error is large at 171 days: the standard error deviation is, therefore, 47% of the average value of the actual days.

4.3 Identification of a Homogeneous Data Set

To analyze this data set unencumbered by the impact of the recognized major productivity differences between different programming workbenches, the data set was divided into two: 33 projects developed on the IBM mainframe and four projects developed on minicomputer platforms and different programming workbenches (IBM S/36 and AS/400 minicomputers).

After an analysis of the 33 mainframe-based projects, 32 of these were classified as major enhancements to existing transaction-based software applications. A single project developed a totally new software application, non-transaction-based, using new technology and by-passing the in-house standard development methodology; furthermore, this new application did not meet user requirements and was never used. This observation was, therefore, considered to be an outlier and discarded.

This reduced data set of 32 observations corresponds closely to the characteristics of a homogeneous development environment as defined in [30], [31]: Many similar projects developed for the same application domain and a standard development process model used over the data collection period. These 32 observations correspond to projects developed in a homogeneous development environment unburdened by major differences in production and quality factors.

To investigate the research questions, additional information was required from the data set: the end results in terms of total FP counts were not sufficient, and the intermediate results of each step of the FPA measurement process were required. Unfortunately, the detailed information on the primary components (data element: DET; logical groups of data elements: RET and FTR) for each of the function types for 11 of the 32 observations was not available, and these 11 projects had to be left out for the purpose of analysing each of the measurement steps. This left 21 observations available for the empirical designs. The average size in terms of number of days for these 21 projects is 332 days, slightly lower than the average of 362 days for the 37 observations.

The reduced data set of 21 projects is illustrated in Fig. 4 and consists of projects within the same order of magnitude: the x-axis is now in FP (min: 39 FP, max: 258 FP) and the actual work-effort is in days on the y-axis (min: 52 days, max: 554 days). The average size in terms of number of days for these 21 projects is 332 days, slightly lower than the average of 362 days for the 37 observations. The visual analysis of Fig. 4 indicates clearly a linear relationship between the independent and the dependent variables. The size of the data set is similar to other samples discussed in the literature, within a range of 15 to 30 projects [35].

4.4 Statistical Analysis of the Data Variables

This data set also has many of the characteristics of the well-behaved data set no. 5 described in [35]; this type of data set does not have the undesirable characteristics of non-normal distributions with strong heteroscedasticity of the variables and resulting in V-shaped data sets when represented on two axes with a dependent and an independent variable (ref. Fig. 5), such as some of the data sets used for analytical purposes in [33], [35].

The regression models reported here are based on the least-squares methods used in [2], [8], [14], [15], [18], [26], [28]. When data sets include outliers it is recommended in [35] to build models based on more robust techniques such as the least-square of balanced errors or the least-square of

TABLE 9
EMPIRICAL DATA SETS: AFP AND UFP MODELS

Data sets	Adjusted Function Points AFP: Standard error	Coef. of variation	R^2	Adj. R^2	Unadjusted Function Points UFP: Standard error	Coef. of variation	R^2	Adj R^2
37 Obs.	171	47	0.47	–	168	46	0.49	–
21 Obs.	67	20	0.81	0.80	67	20	0.81	0.80

inverted relative errors. However, for data sets that are fairly homogeneous and with no outliers problem, such as in the case with the data set reported here, it is known, as reported in [35], that the least-square method performs as well as the other two proposed methods; this was illustrated in [35] with four sets of observations coming from heterogeneous environments (multiple organizations or multiple distinct divisions within organizations), and a set of observations from a single and highly standardized environment.

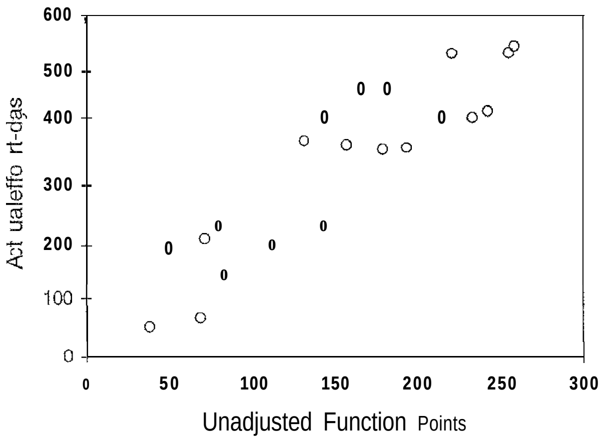


Fig. 4. The empirical data set of 21 observations.

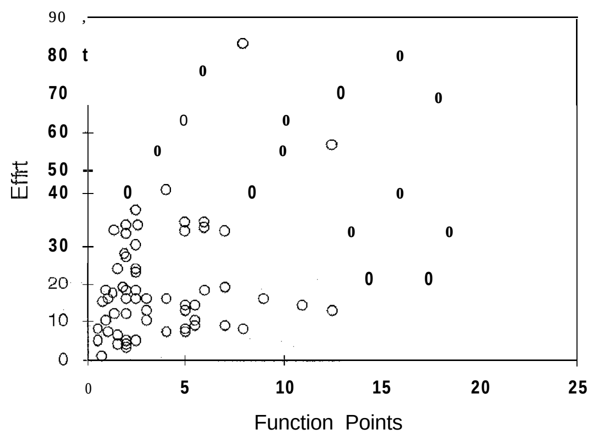


Fig. 5. Data set without normal distribution and with strong heteroscedasticity.

The building of productivity models based on primary components was also discussed in 1331; however, the data set in [33] has many characteristics of heterogeneous data sets and caution must be exercised in the interpretation of the reported results. Furthermore, the analytical perspective was different: In [33], they investigated whether models based on primary components improved the goodness of fit of the models, whereas the research reported here investigate the contribution of primary components in the FPA measurement process and in the building of estimation models.

For the data set used for this research, an analysis was done for the dependent and all independent variables to be used in building the empirical productivity models. The summary statistics of these variables are presented in Table 10: it includes for each variable, the mean, median, minimum and maximum values, skewness and kurtosis factors as well as the Shapiro-Wilk statistics W and Prob. < W for the test of normality. For 20 of the 23 variables, the W value is high and parametric tests such as regression analyzes are appropriate; for only three variables with W values between 0.53 and 0.62 (SDET-D, DET-EF, DET-IP) caution must be exercised when using regression techniques.

With respect to the independent variables, UFP and AFP, the regression models for this set of 21 projects have an R^2 of 0.81 for both the adjusted and the unadjusted FP models (Table 9) and an adjusted R^2 of 0.80. These results confirm the relative accuracy of the graphical analysis of Fig. 4, which exhibits a strictly increasing linear relationship between unadjusted function points (UFP) and work effort.

5 Productivity Models Based on Primary Components

The results of the set of regression models based on the primary components of FPA (e.g., without the weights) are presented next. These productivity models are built using the data from the homogeneous data set of 21 projects from Fig. 4. Independent variables of the models based on the count of primary components (data elements and logical groups of data elements) do not take into consideration any of the transformations and algorithms described in FPA measurement process; these variables are not derived from classification within ranges of values, attribution of a level of complexity from an assignment of weights; they do not require a transformation of values through different types of scale. In fact, the independent variables of this set of models are based strictly on numbers of the absolute scale type.

To analyze the adequacy of each regression model, the regression equation is presented in the tables as well as the standard error of the estimate, the coefficient of variation, the coefficients of multiple determination R^2 and the adjusted R^2. In addition to the models adequacy and models stability, the models aptness concerns were looked at, as recommended in [35]. The models aptness is defined in [35] as the conformity of the residuals to the assumptions that the error values in the regression models are distributed as independent, normal random variables with mean zero and identical variances (normality, independence and homoschedasticity) [34],[35]. Therefore, the statistics W and Prob. < W of the residuals, the Fischer statistics F and p value (F) of the models as well as the Student's statistics T and the (Prob. >|T|) for each of the individual parameters of the models are presented in Appendices B and C.

5.1 Models with Primary Components Only

The two regression models in Table 11 have only one independent variable and are based on either the total number of data elements (TDET) or the total number of groups of data elements (TGRE). The model with TDET as the independent variable does not produce good coefficients (Adj. $R^2 = 0.34$), while the model with TGRE gives much better results (Adj. $R^2 = 0.62$), but they are still noticeably lower than those of the AFP and UFP models (Table 9: Adj. $R^2 = 0.80$). Even then, these two models (with TDET and TGRE) are interesting since they indicate that at the initial stage of the FPA meas-

TABLE 10
STATISTICS OF THE DATA SET VARIABLES (WITH 21 OBSERVATIONS)

*Variable	Mean	Median	Min	Max	Skewness	Kurtosis	Normal W	Normal p(W)
Dependent variable								
Days	332	369	52	544	–0.39	–0.98	0.93	.1557
Independent variables *WITHOUT WEIGHTS*								
TDET	397	329	43	1092	1.17	0.72	0.87	.0081
TGRE	65	56	13	186	1.59	4.45	0.87	.0093
SDET-D	135	94	4	782	3.26	11.88	0.59	.0001
SGRE-D	23	22	3	65	1.12	1.59	0.92	.0896
SDET-TR	262	214	39	988	1.99	5.02	0.82	.0007
SGRE-TR	42	33	10	121	1.33	2.81	0.88	.0146
DET-IF	51	33	0	265	2.26	5.29	0.72	.0001
GRE-IF	6	6	0	20	0.84	0.90	0.92	.0859
DET-EF	84	52	3	517	3.27	12.56	0.62	.0001
GRE-EF	17	15	1	55	1.34	1.83	0.87	.0074
DET-IP	96	42	0	784	3.58	14.2	0.53	.0001
GRE-IP	14	8	0	56	1.76	2.84	0.79	.0003
DET-OP	130	99	0	348	0.67	–0.57	0.93	.1235
GRE-OP	22	17	0	100	2.37	6.43	0.73	.0001
DET-IQ	36	13	0	172	1.81	2.31	0.70	.0001
GRE-IQ	6	4	0	30	1.99	4.88	0.79	.0003
Independent variables *WITH WEIGHTS*								
FPIF	25	24	0	84	1.16	2.10	0.90	.0399
FPEF	44	34	5	112	0.72	-0.29	0.93	.1494
FPIP	28	21	0	107	1.56	2.29	0.84	.0018
FPOP	42	40	0	133	1.04	1.52	0.93	.1383
FPIQ	14	9	0	55	1.40	1.41	0.83	.0013
UFP	154	157	39	258	-0.10	-1.21	0.95	.3294

**Refer to Table 8 for the description of the abbreviated variable names.*

TABLE 11
MODELS WITH PRIMARY COMPONENTS - ONE VARIABLE

Variable	Equation WE: Work-Effort	Std error	Coef. of variation	R^2	Adj R^2
TDET	WE = 0.31 * TDET + 207.8	123	37	0.37	0.34
TGRE	WE = 3.21 * TGRE + 123.7	93	28	0.64	0.62

urement process, there is already a Work-Effort relationship with respect to the primary components, albeit not as good, in a homogeneous environment.

The next model (Table 12) has two independent variables and takes into account both the total of data elements and the total of groups of data. All coefficients of the regression are improved and, with an Adj. R^2 of 0.76, this model is almost as good as the models with the totals of FP (AFP and UFP).

5.2 Models with Primary Components by Function Groups

The next models (Table 13) are based on the segregation of the primary components into two groups, one for the data-type functions, the other for the transaction-type functions. It can be observed that the 2-variable model for the transaction-type group is better (Adj. R^2 = 0.65) than the corresponding 2-variable model for the data-function type (Adj. R^2 = 0.36). However, the 4-variable model is not better (Adj. R^2= 0.74) than the 2-variable model of Table 12 (Adj. R^2= 0.76) based solely on the total number of data elements and groups of elements which does not discriminate between either data- or transaction-type functions.

These models, which segregate the primary components by function groups, do not, on the one hand, significantly improve the reliability of the coefficients of regression but, on the other hand, do not invalidate Albrecht's implicit model of two different groups of functions (data and transaction type).

5.3 Model with Primary Components by the Five Function Types

The model with 10 independent variables (Table 14) takes into account Albrecht's implicit model of five different function types: the 10 variables are the number of data elements and the number of logical groups of data elements in each of the five FPA function types. The results of this model are interesting: the standard error at 68 and the coefficient of variation at 21 are within the range of the AFP and UFP models, while its Adj. R^2 at 0.79 is close to the 0.80 of the former models. This adjusted R^2 of 0.79 indicates that even with 10 independent variables the regression model is still reliable, but that caution should be exercised since they differ from those of the R^2 of 0.90.

TABLE 12
MODEL WITH PRIMARY COMPONENTS - TWO VARIABLES

Variables	Equation WE: Work-Effort	Std error	Coef. of variation	R	Adj R^2
TDET and TGRE	WE = 0.20 TDET + 2.71 TGRE + 74.94	74	22	0.78	0.76

TABLE 13
MODELS WITH DATA-TYPE AND TRANSACTION-TYPE FUNCTION GROUPS

	Equation WE: Work-Effort	Std error	Coef. of variation	R^2	Adj R^2
Data (logical files)	WE = 0.15 SDET-D + 6.01 SGRE-D + 173.44	120	36	0.43	0.36
Transactions	WE = 0.22 SDET-TR + 3.66 SGRE-TR + 121.10	89	27	0.69	0.65
4 variables	WE = 0.13 SDET-D + 3.30 SGRE-D + 0.26 SDET-TR + 2.31 SGRE-TR + 73.27	77	23	0.79	0.74

TABLE 14
MODEL WITH PRIMARY COMPONENTS BY FUNCTION TYPE

	Regression Equation 21 projects	Std error	Coef. of variation	R^2	Adj R^2
10 variables	WE = 1.92 DETIF – 15.43 GREIF – 0.65 DETEF + 4.78 GREEF + 0.08 DETIP + 0.27 GREIP + 0.35 DETOP + 2.17 GREOP + 0.07 DETIQ + 9.69 GREIQ + 55.28	68	21	0.90	0.79

The results of this model are, again, very interesting from several points of view:

1) Compared to the 2-variable model (TDET + TGRE), this model, which differentiates the primary components TDET and TGRE by function type, adds reliability to the coefficients of the regression model.
2) It confirms the validity of Albrecht's five-function-type model, independently of the subsequent steps in the methodology.
3) With respect to the 2- and 4-variable models, based on the two different function groups (Data and Transaction), the five-function-type model is more reliable, and adds information to the measurement process.
4) With respect to the full Function point model (AFP and UFP), the lo-variable model is as reliable for this data set.

It should be noted in the above equation that the 10 dependent variables represent the variable portion of the project costs, or the slope of the regression line, while the constant residual of 55.28 days can be interpreted as representing the fixed costs of a project in this homogenous development environment. It is worth mentioning that this data set derives from an organization with a repeatable project management process and a development life cycle methodology for all development projects estimated at over 60 days of work effort. This project management structure represents a set of management reports and controls not required from smaller projects. The fixed costs associated with the project management process put in place to reduce the uncertainties and risks associated with large projects seems to be accurately reflected by the residual of 55.28 days, or 16.6% of the average effort of 332 days for this sample.

The analysis of the residuals was carried out for each of the previous models with the aid of diagnostics plots as recommended in [35] with the residuals errors and the predicted value on the axes: it confirmed the independence and the homoscedasticity of the residuals (See Appendix B). The analysis of the F and (Prob. < F) statistics for the models and the T and (Prob >|T|) statistics for each varialble of the models (Appendix C) also confirmed the aptness of the models.

6 MODELS BASED ON FPA WEIGHTS

The results of the set of models based on the weights with the mixing of scale types are presented next. The l-variable models based on either the unadjusted or adjusted FP (UFP or AFP) have already been discussed in presenting the empirical design in Section 4 (Table 9). This section will focus on the model with five independent variables based on the FP by function types: internal logical files, external interface files, inputs, outputs, and inquiries.

It was argued in [33] that the constituent parts of FP should be independent and that there should not be any correlation among the internal logical files (FPIF), the external interface files (FPEF), inputs (FPIP), outputs (FPOP), and inquiries (FPIQ). However, when building multiple-variable models it should be verified that one variable is not an exact substitute for another one (e.g., that the correlation coefficient between the two variables is significantly different than 1) rather than testing that they are totally independlent (e.g., correlation = 0). When multiple variables are used to build regression models, the analysis of each of the regression coefficients will indicate how much each of the variables, independently of the other variables, contribute to the explanation of the relationship with the dependent variable.

The sample correlation coefficients between the dependent and independent variables are presented in Table 15: it can be seen that none are close to 1 (or -1) in pair-wise correlation. The five independent variables can then be taken into consideration for building a 5-variable model.

TABLE 15
CORRELATION COEFFICIENTS BETWEEN FP TYPES

	Days	FPIF	FPEF	FPIP	FPOP	FPIQ
Days	1.0					
FPIF	0.536	1.0				
FPEF	0.494	–0.119	1.0			
FPIP	0.506	–0.016	0.739	1.0		
FPOP	0.449	0.064	0.166	0.018	1.0	
FPIQ	0.482	0.541	–0.075	0.055	–0.154	1.0

TABLE 16
MODEL WITH THE FIVE FP TYPES

	Regression Equation WE: Work-Effort	Std error	Coef. of variation	R^2	Adj R^2
5 variables	WE = 2.19 FPIF + 1.70 FPEF + 1.41 FPIP + 1.99 FPOP + 3.64 FPIQ + 26.71	70	21	0.84	0.79

This 5-variable model (Table 16) takes into consideration all the measurement steps of the Albrecht model, with the exception of the last one which adds all types of function points together. The results of this model are interesting: its Adj. R^2 at 0.79 is almost the same as the 0.80 of the AFP and UFP models, and the standard error at 70 and the coefficient of variation at 21 are within the range of the former models. The adjusted R^2 of 0.79 indicates that even with five independent variables, the regression model maintains its reliability.

This model is particularly interesting for several reasons:

1) Because it includes all the transformations of FPA measurement methodology with respect to range intervals, complexity classification schema and assignments of weights, it does not produce better coefficients of regression than the lo-variable model based solely on the primary components.
2) This model does not, on the other hand, invalidate any of the transformations of FPA measurement methodology with respect to range intervals, complexity classification schema and assignments of weights.
3) Albrecht's 1983 weights are valid in this subset and reference context. This also tends to confirm that the scale transformations, which are not valid when viewed only from a mathematical standpoint, are carried out through implicit valid transformations.
4) The transformations and weights do not add information for this data set from a homogeneous development environment.
5) The last step in the FPA measurement methodology which adds all function-type points together to reduce them to a single number, UFP or AFP, does not improve significantly the regression coefficients (Table 9: Adj. $R^2 = 0.80$; Table 15: Adj. $R^2 = 0.79$). In fact, they are slightly increased, albeit not by much.
6) The results of the full FPA model (Table 9), based on the sets of weights determined by debate and trial on the Albrecht 79 data set, are nevertheless surprisingly reliable. This would lead to the belief that they could be fairly robust with respect to both the transformations of scale types and implicit models.

The analysis of the residuals was carried out for each of the previous models with the aid of diagnostics plots as recommended in [34],[35],[36] which confirmed the independence and the homoscedasticity of the residuals (Appendix B). The analysis of the statistical tests for the F and p(F) tests for the models and the T and p(T) tests for each variable of the models (Appendix C) also confirmed the aptness of the models.

8 SUMMARY AND CONCLUSIONS

The various models and statistical tests utilized have probed the impact of each of the measurement steps of FPA with respect to the work-effort relationship in the reference context of a set of 21 projects in a homogeneous development environment characterized as follows: a mainframe workbench, MIS applications, transaction-based applications, major enhancement projects to existing applications and a defined and stable development process. This type of empirical designs represents a significant contribution to the study of FPA: It allows an analysis of the internal structure of itsmeasurement process. This type of empirical design allows to focus on a single dependent variable, the work-effort, but through multiple models built for each of the sequential steps of the FPA measurement process, both prior and after the mixing of scales.

The analytical and empirical results derived from this type of empirical designs are summarized below:

1) In a homogeneous environment, the 2-variable model based only on the count of the primary components (Table 12) compares favourably with the full FPA model (Table 9). For the data set under scrutiny, this model confirms the existence of an implicit relationship between the primary components and work-effort.
2) In a homogeneous environment, the model based on both the count of the primary components and their grouping in the two function groups (Table 13, 4-variable model: data- and transaction type functions) is almost equivalent to the full FPA model.
3) In a homogeneous environment, the model based on both the count of the primary components and their grouping by the five-function-type implicit FPA

model (Table 14, lo-variable model) is almost equivalent to the full FPA model. This model confirms the existence and reliability of the implicit relationship in the five-function-type FPA model.

4) None of the implicit models and transformations present in the changes of measurement scales and weights (Tables 9 and 16) adds significant information when compared to the results of the model based on the primary components by function-type (Table 14).

5) By the same token, it confirms the robustness of the FPA measurement process with its set of different function-type structure and weights: when the details on the primary components are not known, or not kept in a historical database, the information on FP by the five-function-type FPA productivity model can be used for productivity and estimation analysis.

   It should be noted, however, that the historical database utilized for this research has many of the same characteristics of the Albrecht database of projects. This limits the generalisation of this interpretation; further research1 would be required to verify the pertinence, and robustness, of the weights within homogeneous environments significantly different from the database utilized in this study.

6) The last step in the FPA measurement process, which adds the points of all five function types together (Table 9) adds little to the relationship with work effort.

7) However, for comparison purposes across projects and organizations, the single end number of the FP measurement process is still very useful in reducing the results back to a single (and dimensionless!) number: its simplifies the communication process from a management perspective.

Based on this analysis of FPA, it is recommended to keep all the levels of the detailed counts in the FP historical databases: this will enable more accurate analyzes without the burden of the effects of the implicit models and transformations. This recommendation also applies to software vendors who are currently working on automated FP counters.

This research project confirms the benefits, as mentioned in [32], of storing the basic values produced by the measurement process of algorithmic metrics such as FPA: it will facilitate the control of the measurement process as well as more in-depth analysis of the relationship(s) under scrutiny.

In conclusionthis research work illustrates the impact of the implicit models, and their usefulness, of each step of the FPA measurement process for the study of the work-effort relationship. These results can be significant for both productivity analysis and for estimating using FP. Furthermore, this research work provides a methodology for in-depth studies of algorithmic metrics utilized in productivity and estimation models.

APPENDIX A — DETAILED DATA SET OF 21 OBSERVATIONS

TABLE Al
Variables

VARIABLE	DESCRIPTION
	Upper part of the table
OBS	Sequential number of the observation in the R&D Database
IDAP	Identification number of the observation given by the organization
LEF	Number of Data Element Type (DET) for the External Logical Files
NEF	Number of Record Element Type (RET) for the External Logical Files
PEF	Number of points for the External Logical Files
LIF	Number of Data Element Type (DET) for the Internal Logical Files
NIF	Number of Record Element Type (RET) for the Internal Logical Files
PIF	Number of points for the Internal Logical Files
LIP	Number of Data Element Type (DET) for the Inputs
NIP	Number of File Type Referenced (FTR) for the Inputs
PIP	Number of points of the Inputs
LOP	Number of Data Element Type (DET) for the Outputs
NOP	Number of File Type Referenced (FTR) for the Outputs
POP	Number of points of the Outputs
LIQ1	Number of Data Element Type (DET) of the input part of the Inquiries
NIQ1	Number of File Type Referenced (FTR) by the input part of the Inquiries
LIQ2	Number Data Element Type (DET) of the ouptut part of the Inquiries
	Bottom part of the table
OBS	Sequential number of the observation in the R&D Database
NIQ2	Number of File Type Referenced (FTR) by the ouptut part of the Inquiries
PIQ	Number of points of the Inquiries
UFP	Unadjusted Function points
GSC	Value adjustment Factor
AFP	Adjusted Function points
ACD	Number of Actual Days
LIQ	Selected number of the DET for the Inquiries
NIQ	Selected number of FTR for the Inquiries
TELE	Number of DET
TENT	Number of FTR/RET
REF	FP ratio for the Internal Logical Files
RIF	FP ratio for the External Interface Files
RIP	FP ratio for the Inputs
ROP	FP ratio for the Outputs
RIQ	FP ratio for the Inquiries

TABLE A2
Data

OBS	IDAP	LEF	NEF	PEF	LIF	NIF	PIF	LIP	NIP	PIP	LOP	NOP	POP	LIQ1	NIQ1	LIQ2
1	1	52	17	69	32	12	38	251	56	107	0	0	0	6	4	34
2	2	138	46	112	0	0	0	0	0	0	0		0	13	13	172
3	3	143	16	75	40	6	24	42	21	36	99	17	35	10	2	11
4	4	111	27	84	57	11	42	105	24	41	133	17	45	24	9	119
5	5	24	9	17	70	11	35	84	15	36	178	42	71	25	5	66
6	6	12	1	5	92	12	84	784	51	81	204	16	51	0	0	0
7	7	33	2	25	4	1	7	29	15	18	40	12	23	3	3	11
8	8	103	8	57	34	4	17	57	8	18	348	17	47	1	1	18
9	9	5	2	5	22	6	21	44	12	24	68	7	23	2	2	25
10	11	517	18	67	265	5	29	90	11	21	66	22	40	0	0	0
11	24	49	14	29	79	8	31	78	17	51	320	24	83	2	1	64
12	25	176	16	45	199	20	61	215	28	63	219	17	59	8	2	160
13	27	78	13	34	6	1	7	7	7	21	266	16	31	4	3	17
14	28	87	55	99	11	10	35	13	7	17	170	100	84	31	14	10
15	29	3	2	10	1	1	7	3	3	9	33	5	7	0	0	3
16	33	44	17	30	57	5	27	112	8	24	116	25	64	0	0	0
17	35	28	15	30	0	0	0	17	1	4	50	9	17	0	0	0
18	36	59	31	30	33	7	17	41	1	4	58	10	18	3	1	4
19	37	65	28	55	0	0	0	2	2	6	246	64	133	0	0	0
20	43	18	12	35	15	4	21	12	2	6	98	29	45	2	2	5
21	44	21	4	15	47	7	31	24	4	6	26	5	10	2	2	13

OBS	NIQ2	PIQ	UFP	GSC	AFP	ACD	LIQ	NIQ	TELE	TENT	REF	RIF	RIP	ROP	RIQ	
1	11	19	233	87	203	418	34	11	369	96	30	16	46	0	8	
2	30	55	167	79	132	468	172	30	310	76	67	0	0	0	33	
3	6	9	179	80	143	360	11	6	335	66	42	13	20	20	5	
4	16	43	255	80	204	531	119	16	525	95	33	16	16	18	17	
5	9	22	181	80	145	471	66	9	422	86	9	19	20	39	12	
6	0	0	221	85	188	525	0	0	1092	80	2	38	37	23	0	
7	5	7	80	80	64	225	11	5	117	35	31	9	23	29	9	
8	2	5	144	79	114	229	18	2	560	39	40	12	13	33	3	
9	4	10	83	87	72	143	25	4	164	31	6	25	29	28	12	
10	0	0	157	86	135	369	0	0	938	56	43	18	13	25	0	
11	7	22	216	66	143	416	64	7	590	70	13	14	24	38	10	
12	4	14	242	72	174	428	160	4	969	85	19	25	26	24	6	
13	11	39	132	78	103	377	17	11	374	48	26	5	16	23	30	
14	5	23	258	90	232	544	31	14	312	186	38	14	7	33	9	
15	2	6	39	80	31	52	3	2	43	13	26	18	23	18	15	
16	0	0	145	75	109	400	0	0	329	55	21	19	17	44	0	
17	0	0	51	80	41	187	0	0	95	25	59	0	8	33	0	
18	2	3	72	89	64	198	4	2	195	51	42	24	6	25	4	
19	0	0	194	74	144	363	0	0	313	94	28	0	3	69	0	
20	2	9	113	70	79	195	5	2	148	49	31	19	5	40	5	
21	4	7	69	78	54	69	13	4	131	24	22	45	9	14	10	

APPENDIX B – ANALYSIS OF THE RESIDUALS OF THE LINEAR REGRESSION RESULTS

TABLE B1
MODELS BASED ON THE PRIMARY COMPONENTS (WITHOUT THE WEIGHTS)

Model and variables	RESIDUALS	
	W	Prob. < W
1 var: TDET	0.97	.6928
1 var.: TGRE	0.96	.4169
2 var: TDETTGRE	0.94	.1674
2 var.: SDET-D, SGRE-D	0.99	.99
2 var.: SDET-TR, SGRE-TR.	0.96	.43
4 var.	0.93	.1587
10 var.	0.91	.0590

TABLE B2
MODELS BASED ON THE UFP (WITH THE WEIGHTS)

Model and variables	RESIDUALS	
	W	Prob. < W
1 var: UFP	0.94	.1949
5 var.	0.94	.2240

APPENDIX C – LINEAR REGRESSION RESULTS

TABLE CI
MODELS BASED ON THE PRIMARY COMPONENTS (WITHOUT THE WEIGHTS)

Model and variables	MODEL				VARIABLES			
	R^2	Adj. R^2	F	Prob. < F	Variable	Coef.	T	Prob. < T
1 var: TDET	0.37	0.34	11.26	.0033	TDET	0.31	3.356	.0033
					Constant	207.82	4.55	.0002
1 **var.: TGRE**	**0.64**	**0.62**	33.55	.0001	TGRE	3.21	5.79	.0001
					Constant	123.67	2.99	.0074
2 **var:** TDET **and** TGRE	**0.78**	**0.76**	**32.40**	.0001	TDET	0.20	3.45	.0028
					TGRE	2.71	5.89	.0001
					Constant	74.94	2.09	.0507
2 **var.:** SDET-D **and** SGRE-D	**0.43**	**0.36**	**6.74**	.0065	SDET-D	0.15	0.89	.3819
					SGRE-D	6.01	3.31	.0039
					constant	173.44	3.43	.0030
2 **var.:** SDET-TR **and** SGRE-TR	**0.69**	**0.66**	**19.88**	.0001	SDET-TR	0.22	2.23	.0390
					SGRE-TR	3.65	4.37	.0004
					constant	121.10	3.13	.0057
4 **var.**	**0.79**	**0.73**	15.19	.0001	SDET-TR	0.26	2.85	.0116
					SGRE-TR	2.31	2.37	.0306
					SDET-D	0.13	1.23	.2354
					SGRE-D	3.29	2.11	.0510
					constant	73.27	1.95	.0689
10 **var.**	**0.90**	**0.79**	**a.72**	0.0010	DET-IF	1.92	1.58	.1456
					GRE-IF	–15.43	-1.69	.1220
					DET-EF	–0.65	-1.08	.3037
					GRE-EF	4.78	1.81	.0995
					DET-IP	0.08	0.39	.7048
					GRE-IP	6.27	2.37	.0391
					DET-OP	0.35	1.83	.0966
					GRE-OP	2.17	1.85	.0945
					DET-IQ	0.08	0.10	.9215
					GRE-IQ	9.69	0.79	.1031
					constant	55.26	1.30	.1060

TABLE C2
MODELS BASED ON THE UFP (WITH THE WEIGHTS)

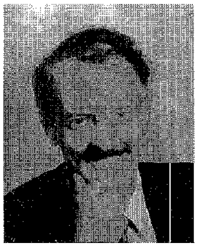
Model and Variables	MODEL				VARIABLES			
	R^2	Adj. R^2	F	Prob. < F	Variable	Coef.	T	Prob. < T
1 var: UFP	0.81	0.80	83.48	.0001	UFP	1.95	9.14	.0001
					constant	30.85	0.86	.4023
5 var.	.84	.78	15.48	.0001	FPIF	2.17	1.84	.0850
					FPEF	1.72	2.75	.0149
					FPIP	1.44	1.67	.1162
					FPOP	1.99	3.91	.0014
					FPIQ	3.59	2.88	.0114
					constant	26.00	0.67	.5142

ACKNOWLEDGMENTS

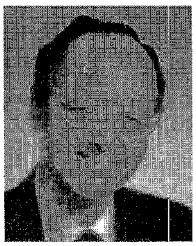
The authors would like to thank the anonymous reviewers for their comments, more specifically, about the statistical analysis of this data set. This work has been supported by Bell Canada. The opinions expressed in this article are solely those of the authors.

REFERENCES

[1] A.J. Albrecht, "Measuring Application Development Productivity," *Proc. IBM Applications Development Symp.*, Monterey, Calif., Oct. 14-17, 1979.
[2] A.J. Albrecht and. J.E. Gaffney, "Software Functions, Source Lines of Code, and Development Effort Prediction: A Software Science Validation," *IEEE Trans. Software Eng.*, vol. 9, no. 6, Nov. 1983.
[3] IBM CISGuidelines 313, *AD/M Productivity Measurement and Estimate Validation, Nov. 1984.*
[4] International Function Point Users Group (IFPUG), *Function Point Counting Practices Manual, Release 1.0*, IFPUG, Westerville, Ohio, 1986.
[5] International Function Point Users Group (IFPUG), *Function Point Counting Practices Manual, Release 2.0*, IFPUG, Westerville, Ohio, 1988.
[6] International Function Point Users Group (IFPUG), *Function Point Counting Practices Manual, Release 3.0*, IFPUG, Westerville, Ohio, Apr. 1990.
[7] International Function Point Users Group (IFPUG), *Function Point Counting Practices Manual, Release 4.0*, IFPUG, Westerville, Ohio, Jan. 1994.
[8] C.F. Kemerer, "An Empirical Validation of Software Cost Estimation Models," Comm. *ACM*, vol. 30, no. 5, May *1987*.
[9] C.F. Kemerer, "Function Point Measurement Reliability: A Field Experiment," *Proc. IFPUG Fall Conf.*, San Antonio, Tex., Oct. 1990.
[10] G.C. Low and D.R. Jeffery, "Function Points in the Estimation and Evaluation of the Software Process," *IEEE Trans. Software Eng.*, vol. 16, no. 1, Jan. 1990.
[11] E.E. Rudolph, *Productivity* in *Computer Application Development*, Dept. of Management Studies, Univ. of Auckland, New Zealand, 1983.
[12] E.E. Rudolph, "Precision of Function Point Counts," *Proc. IFPUG 2989 Spring Conf.*, San Diego, Calif., Apr. 1989.
[13] C.A. Behrens, "Measuring the Productivity of Computer Systems Development Activities with Function Points," *IEEE Trans. Software Eng.*, vol. 9, no. 6, Nov. 1989.
[14] R.D. Emrick, "Software Development Productivity-Second Industry Survey," *Proc. IFPUG 1988 Spring Conf.*, Dallas, May 1988.
[15] R.D. Emrick, "Further Analysis-Software Development Productivity-second Industry Survey," *Proc. IFPUG Fall Conf.*, Montreal, Sept. 1988.
[16] J.E. Gaffney, "The Impact on Software Development Costs of Using HOL's," *IEEE Trans. Software Eng.*, vol. 12, no. 3, Mar. 1986.
[17] J.-M. Desharnais, *Analyse sfatistique de la productivité des projets de développement en informafique à partir de la technique des points de fonction,* Maitrise en informatique de gestion, Université du Quebec à Montreal, Dec. 1988.
[18] J.-M. Desharnais, "Adjustment Model for Function Point Scope Factors-A Statistical Study," *Proc. IFPUG Spring Conf.*, Fla., Apr. 1990.
1191 C. Jones, and Capers, "Feature Points (Function Point Logic for Real Time and System Software)," *Proc. IFPUG Fall 1988 Conf.*, Montreal, Quebec, Oct. 1988.
[20] C.R. Symons, "Function Points Analysis: Difficulties and Improvements, *IEEE Trans. Software Eng.*, vol. 14, no. 1, Jan. 1988.
[21] N.E. Fenton, *Software Metrics-A Rigorous Approach,* London: Chapman Hall, 1991.
[22] H. Zuse, "Software Metrics-Methods to Investigate and Evaluate Software Complexity Measures," *Proc. Second Ann. Oregon Workshop on Software Metrics,* Portland, Oreg., Mar. 19, 1990.
[23] H. Zuse, *Software Complexity-Measure and Methods,* Walter de Gruyter, ed., Berlin-New York, 1991.
[24] A. Abran and P.N. Robillard "Identification of the Structural Weakness of Function Point Metrics," *Proc. Third Ann. Oregon Workshop on Software Metrics*, Portland, Mar. 18, 1991.
[25] A. Abran, "Function Points: A Study of Their Measurement Processes and Scale Transformations," *J. Systems Software,* vol. *25*, pp. 171-184, 1994.
1261 A. Abran, "Analyse du processus de mesure des points de fonction," PhD. thesis, École Polytechnique de Montreal, Montreal, Quebec, Canada, Mar. 1994.
[27] L.O. Eijogu, Software *Engineering with Formal Metrics.* Wellesley, Mass.: QED Information Sciences, 1991.
[28] R.D. Banker and C.F. Kemerer, "Scale Economies in New Software Development," *IEEE Trans. Software* Eng., vol. 15, no. 10, Oct. 1989.
[29] F.J. Knaff and J. Sacks, "Software Development Effort Prediction Based on Function Points," *Proc. COMPSAC '86*, Ill., 1986.
[30] S.G. MacDonell, "Rigor in Software Complexity Measurement Experimentation," *J.* Systems Software, Oct. 1991.
[31] C.L. Ramsey and V.R. Basili, "An Evaluation of Expert Systems for Software Engineering Management," *IEEE Trans. Software Eng.*, vol. 15, no. 6, pp. 747-759, June 1989.
1321 B. Kitchenham and C. Jaggers, "State of the Art Survey-vol. 2: Software Metrics," Esprit project P2046: MERMAID, Feb. 24, 1989.
1331 B. Kitchenham and K. Kansala, "Inter-Item Correlations Among Function Points," *Proc. First Int'l Software Metrics Symp.*, IEEE CSPress, Los Alamitos, Calif., pp. 11-14, May-21-22, 1993.
[34] Y. Miyazaki, M. Terakado, K. Ozaki, and H. Nozaki, "Robust Regression for Developing Software Estimation Models," *J. Systems Software,* vol. 27, pp. 3-16, 1994.
[35] J.E. Matson, B.E. Barrett, and J.M. Mellichamp, "Software Development Cost Estimation Using Function Points," *IEEE Trans. Software Eng.*, vol. 20, no. 4, pp. 275-287, Apr. 1994,.
[36] R.D. Banker, S.M. Datar, C.F. Kemerer, and D. Zweig, "Software Complexity and Maintenance Costs," *Comm. ACM*, vol. 36, no. 11, pp. 81-94, Nov. 1993,.



Alain Abran holds masters degrees in management sciences, 1974, and electrical engineering, 1975, both from the University of Ottawa, and a PhD degree in software engineering, 1994. from École Polytechnique de Montreal. Canada. His research interests include software productivity and estimation models, risk management, functional size measurement models, and econometrics models of software reuse. Dr. Abran is the director of the research laboratory in software engineering management and a professor at the Universite where he teaches graduate courses in software engineering. Dr. Abran has more than 20 years of industry experience in information systems development and software engineering. The maintenance measurement program he developed and implemented at Montreal Trust, Canada, received one of the 1993 *Best of the Best* awards from the Quality Assurance Institute, Orlando, Florida. He is a member of the IEEE



Pierre N. Robillard received the PhD degree from Universite Laval, Québec, Canada, in 1977. He is a full professor of software engineering in the Ecole Polytechnique, Montreal, Canada. He leads the Software Engineering Research Laboratory, where graduate students are doing research on software modeling, software design, software process adaptation, software process measurements, and quality control and assessment. **His interests include cognition as applied to software design and specification, software** process evaluation and adaptation, quality assessment, and software metrics, He has been involved in numerous commercial CASE tool projects, including structured base editor, software quality evaluation tool, and simulation engine. He is currently adapting software process based on rationale and cognitive science. Dr. Robillard is co-chair of the National Council of Computer Program Accreditation. He is a member of the board of directors of organizations that-specialize in technological transfer of software engineering tools and methods. He is a member of the IEEE.