ISSN:2231-0711

Geetika Vyas et al | IJCSET(www.ijcset.net) | February 2016 | Vol 6, Issue 2,82-88

Empirical Evaluation of Metrics to Assess Software Product Line Feature Model Usability

Geetika Vyas[#1] and Dr.Amita Sharma[#2]

[#1, #2] *The IIS University, Jaipur Rajasthan, India*

***Abstract*-A Software product line represents systems having conceptual similarity. All the systems in the product line have commonalities and variability. Feature models are often used to represent this intrinsic commonality and variability. They have a tree-like structure. A feature model which is low in quality will have negative effect over all the products belonging to the product line. Thus, early indicators in the form of metrics are required to assess quality of feature models. Assessments of quality attributes will help in avoiding the consequences of inferior quality and faulty design at the later stages of production. Quality attributes are of two types: first type of attributes is internal which can be measured through product related features like length, complexity, efficiency etc. Second are external attributes which can be assessed once the product is fully functional like usability, reliability, maintainability etc. Usability is an external attribute which focuses on easy and efficient product usage. In reference to feature models usability is an essential quality to be possessed. In this paper, we intend to validate an existing structural metrics for software product line feature models. In our research, we try to examine whether the available metrics are fair indicators of the three main sub characteristics of usability viz. learnability, understandability and communicativeness. We try to analysis whether these existing metrics for feature models have a correlation with feature model usability. For this we have employed statistical correlation techniques. Results obtained from the empirical validation shows that the metrics are correlated to the subjective perception regarding the usability of the feature models.**

***Keywords*-Software product lines, feature models, structural complexity metrics, empirical validation.**

1. INTRODUCTION:

Software Product Lines (SPL) Engineering develops and maintains families of products keeping track of their commonality and variability [1]. This paradigm owns many advantages like: reduced time-to-market and time-to revenue, improved competitive product value, quality of the product and company reputation, better scalability of business models and agility, reduction in product deployment risk etc. It consists of two processes namely domain engineering (also called engineering-for-reuse) and application engineering (engineering-with-reuse) [2]. Feature models are often used to represent the intrinsic commonality and variability found in software product lines. These models can be easily extended and modified in order to meet the user requirements. Feature models were first introduced by Kang et al and are often used to diagrammatically represent the features available in a product line. They portray all the configurations that a product line can possibly have [3]. The concept of feature is useful not only during the analysis and design phase but also implementation of software product lines.

Quality attributes have a significant influence on software product lines. They help in analyzing and evaluating the quality of the software product line. Usability is considered as one of the most important quality factors in reference to SPL. It is a user quality and can be defined in terms of ease of use [4]. Feature models must be user friendly. They should be easy to learn and operate, with simple navigation and interpretation of output for all types of users. Usability defines how well the feature model meets the user requirements i.e. the variability and commonality [5]. The higher the degree of usability experienced, the larger is the understanding of the product line. Usability is concerned with evaluating how well the model is understandable and communicative. It also affects reusability of feature models. The level of usability acts as a major determinant of the success or failure of a product line. Usability assessment will thus help designers make design improvements in the product line leading to more usable product lines. The challenge of developing more usable product lines has promoted the research for assessing the usability of feature models.

Although attempts have been done in the domains of software measurement for enhancing product quality, but most of them practice the goal of evaluations in later stages by using measurements which are quantitative by nature. Whereas, to develop a better product line the quality characteristic of early artifacts should be measured. Assessing artifacts like feature models early in the life-cycle will help in obtaining a prediction model for quality characteristics like usability. Feature models amount to be an important object while developing software product lines. Their quality is vital as this will have immense impact on the quality of all the products of the SPL. After reviewing the existing measures that can be applied to feature models we have come across a set of measures for SPL feature models proposed by Bagheri et al. (2011)[6]. Our theme is to exercise over these measures and predict feature models usability early in the product line development. We carried out a controlled experiment to evaluate if there is empirical evidence that SPL feature model structural complexity metrics is correlated with usability sub-characteristics: such as Learnability, understandability, and communicativeness. In a nutshell the major contributions of this paper are:

a) To describe the benefits of assessing usability quality attribute in reference to SPL feature models

b) To empirically validate the existing metrics to assess feature model usability

c) To evaluate the level of correlation between the metrics and usability of Feature Models

The remaining paper is organized as follows: quality and its various attributes are introduced in Section II; Section III contains literature review over structural metrics. The experimental setup and design is described in Section IV. Results and analysis is presented in Section V. General discussions and concluding remarks are presented in section VI and VII. The paper is concluded in section 8.

2. QUALITY ATTRIBUTES:

The quality of software systems heavily depends over the accurateness of the requirements specification [7]. Thus the focal point should be to introduce quality early during development of the product line. Feature models graphically represent all the product line features. They are a basic source of requirement specifications and they lay the foundation for the product line family. For this reason, their quality significantly impacts the quality of all the systems in the product line family. Thus enhancing the quality of feature diagrams will prove to be a key footstep towards enhancing the product line quality.

Quality is a multidimensional notion, which comprises of variety of characteristics such as functionality, reliability, usability, efficiency, maintainability and portability [ISO 99]. Usability addresses product quality from the perspective of its users, i.e. how easy the users find it to use or learn [8]. ISO/IEC 9126-1 mentions usability as an important quality characteristic which should be considered during the development phase.

The literature also offers a variety of definitions for Usability:

(a) The capability of the software product to be understood learned, used and attractive to the user, when used under specified conditions (ISO/IEC 9126-1:2000).

(b) The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use (ISO 9241-11:1998).

(c) The ease with which a user can learn to operate, prepares input for, and interprets output of a system or component (IEEE Std.610.12-1990).

(d) Usability of a software product is the extent to which the product is convenient and practical to use [8].

(e) The probability that the operator of a system will not experience a user interface problem during a given period of operation under a given operational profile [11].

Observing the potential future importance of usability standards, it seems reasonable enough to study and analyze measures to assess usability quality attribute in reference to feature models applicable for SPL. This assessment will lead to increased usability eventually leading to increased productivity, usage, adoption, support, user satisfaction and reduced development time and costs. Usability can be evaluated by the quality of communication or interaction between the system and its user. The unit of measurement can be the user's behavior i.e. level of satisfaction, comfort, time spent in performing an action, etc. in natural, virtual or physical environment where communication between user and product occurs.

3. STRUCTURAL METRICS:

Metrics provide an important way towards improving the software quality. They can be used to understand, control, and improve product development phase. Metrics are categorized as code and structure based. Structural metrics assess the physical composition and configuration of the system. This makes these metrics early indicators of product quality. Literature review reveals that several metrics are proposed but limited to the domain of object oriented systems, UML diagrams, and program code [9][10][11][12] [13][14].

With the advent of Feature Oriented Programming, feature models were introduced. And a need for assessing the structural complexity of the feature models was realized. Unfortunately, there is a little reference about structural metrics for feature models in the existing literature. The first approach towards the definition of metrics for feature models was found in the work of Bagheri et al. The authors have proposed a set of metrics to measure the structural complexity of feature models. It was observed that structural complexity of a feature diagram is influenced by various elements that compose it. The metrics proposed by the authors are: Size Measures:

a) Number of Features (NF)
b) Number of Top Features (NTop)
c) Number of Leaf Features (NLeaf).

Structural Complexity Measures:

a) Cyclomatic Complexity (CC)
b) Cross-Tree Constraints (CTC)
c) Ratio of Variability (RoV)
d) Coefficient of Connectivity Density (CoC)
e) Flexibility of Configuration (FoC)
f) Number of Valid Configurations (NVC).

Length Measure:

a) Depth of Tree (DT)

In their experiment Bagheri et al have proposed structural metrics to assess SPL feature models maintainability. The focal point of our research is to use these metrics and assess the usability of SPL feature models. Usability being a very important quality attribute in the domain of SPL should be assessed in the early phases of development. Feature models are used to depict variability and commonality in the early phases of development in SPL. And only useful feature models will lead the product line to gain its objectives of reduced cost, time and effort. Therefore, in our research we aim to assess usability of the feature models. Also, as research in this area is very scarce, further experimentation will help in setting standards for assessment of quality attributes. This research seems to be the need of the hour.

4. EXPERIMENTAL DESIGN AND SETUP

A. *Goal of the study*

For successful conduction of any experiment, the goals should be clearly identified and specified. We have used the standardized Goal-question-Metric (GQM)[15] template. This helped us in clearly identifying and specifying the goal of our experiment. The goal thus set is shown in table1.

TABLE 1: GOAL OF THE EXPERIMENT TO BE CONDUCTED.

Analyze	*Structural complexity metrics for SPL feature models*
For the purpose of	*Evaluating*
From the point of view of	*Researchers*
In the context of	*MCA semester V students*

B. Variables

To conduct experiments, hypothesis is needed along with a set of variables. They help in measuring the experiment and during the analysis phase.

Independent Variables:

In our case independent variables are the structural metrics. They are categorized as independent because within the cause-effect relationship which is of our major concern, they represent the cause, i.e. we want to study if these metrics are or aren't correlated with usability (and its sub characteristics) of software product line feature models.

Dependent Variables

Because external quality attributes represent the effect in the cause-effect relationship, in our experiment the dependent variable is feature model usability.

C. Hypotheses formulation

The foundation of empirical studies is the definition and validation of hypothesis. Our intention is to study whether the structural metrics have a meaningful association with attributes like usability. According to ISO 9126 usability is a vital external quality attribute. It can be easily explained in terms of its three sub characteristics viz. Learnability, understandability, and communicativeness. The aim of our experiment is to analyze whether the structural measures proposed for feature models are suitably serving as indicators for the evaluation of usability. This leads us to the formulation of these hypotheses:

Null hypothesis H0:

There is no significant correlation between the structural metrics and usability of SPL feature models.

Alternative hypothesis H1:

There is a significant correlation between the structural metrics and usability of SPL feature models.

D. Objects of Study:

The feature models included in our experiment are picked from Software Product Line Online Tools (SPLOT). It is a publicly available online repository consisting of feature models. All feature models were validated and checked for possible dead features. While selecting the models different domains were kept in mind making them suitable for the experiment. Total 13 feature models were selected keeping in mind their understandability by the subjects of the study. The language for the models was restricted to English only.

E. Data Collection

The aim of this study is to recognize relationship among the existing metrics with the subjective perception about the usability of feature models. In future this identification will support prediction of feature model usability. The values for the independent variables were undemanding and were calculated as metrics values for all the feature models. All the metric values thus obtained are tabulated in table 2:

TABLE 2: METRIC VALUES FOR ALL MODELS INCLUDED IN THE EXPERIMENT

Feature Model	NF	NTop	Nleaf	CC	CTC	RoV	CoC	FoC	NVC	DT
1	12	4	8	0	0.429	2.333	0.917	0.000	28	3
2	10	4	7	1	2.000	2.500	0.900	0.100	20	3
3	12	2	8	0	5.000	2.250	0.917	0.083	60	4
4	15	3	11	2	3.800	2.200	0.933	0.200	57	4
5	12	5	8	0	2.000	2.000	0.917	0.250	24	4
6	10	4	7	2	1.400	2.500	0.900	0.200	14	3
7	17	3	12	0	7.765	3.250	0.941	0.411	132	5
8	10	2	7	0	0.200	3.500	0.900	0.100	2	3
9	18	5	11	2	0.722	2.000	0.944	0.277	13	5
10	11	3	6	2	0.909	1.750	0.909	0.363	10	4
11	17	6	13	0	8.471	3.333	0.941	0.470	144	4
12	14	4	10	2	9.000	3.000	0.929	0.357	126	4
13	27	9	20	2	42.667	2.833	0.963	0.222	1152	4
Mean	14.231	4.154	9.846	1.000	6.489	2.573	0.924	0.233	137.077	3.846
Std. Dev.	4.764	1.864	3.760	1.000	11.317	0.562	0.020	0.140	309.024	0.689

The subjective perception of the participants was obtained through questionnaires. The process to gather the subjective perception was as follows: the participants had taken a course in software engineering; they were given a demo class. The participants were kept unaware about the aspects and hypothesis of the study. They were given time to communicate their queries about the models and their semantics. After this they were given the questionnaires to assess their subjective perception. In the questionnaire, each question consist a set of 3 sub questions (one for each sub characteristic) for all 13 models. The questions queried the level of usability of the models on the basis of 7 point likert scale [16] as shown in table 3.

TABLE 3: LINGUISTIC VALUES FOR THE SUBJECTIVE EVALUATION OF THE SUB CHARACTERISTICS OF USABILITY.

Extremely Difficult	Very Difficult	A Bit Difficult	Neither Difficult Nor Easy	Quite Easy	Very Easy	Extremely Easy
1	2	3	4	5	6	7

Some 140 subjects participated in the experiment. The subjective perception of all the participants is shown in table 4. The values are based on the linguistic values that were shown in table 3. The values in table 4 are the median values of the opinion referring to each sub characteristic of usability

.

TABLE 4: SUBJECTIVE OPINION OF THE PARTICIPANTS ABOUT FEATURE MODEL USABILITY

Feature Model	Communicativeness	Understandability	Learnability
Car	6	6	6
E-Shop	5	6	5
Watch Model	5	5	6
Graph Product Line	5	5	5
ATM	5	5	5
Mobile Phone	6	6	6
Direct To Home	5	5	5
Computer System	6	7	7
Online Examination Product Line	5	5	5
Smart Home	5	5	5
Text Editor	5	5	5
Search Engine	5	5	5
Bicycle	5	6	6

F. Validation of data

Once the data was collected, to ascertain the degree of consensus among the subjects we employed the Cronbach's Alpha [17]. This analysis is significant as the subjects should reach a certain level of agreement else convincing conclusions cannot be drawn. Therefore we used the Cronbach's Alpha to retrieve the level of similarity among the qualitative behavior of the participants. Results obtained from the test are shown in table 5.

TABLE 5: CRONBACH'S ALPHA FOR DEGREE OF RESEMBLANCE BETWEEN THE OPINIONS OF THE PARTICIPANTS

Cronbach's Alpha			
No of Items	Communicativeness	Understandability	Learnability
13	.900	.894	.903

As seen in the above table, the degree of reliability of all the participants for all the three sub characteristics is higher than .7. It indicates that there exists a reasonable agreement between the participants. In other words there is a fair resemblance or homogeneity between the opinions of all the participants. As a result of this reliability analysis, we conclude that the opinion of the participants is reliable enough for further analysis.

5. ANALYSIS TECHNIQUES

All the thirteen feature models employed in our experiment are from diverse domain and thus form satisfactory set of objects for the study. They also vary in the metric values. The data collected empirically is also quantitatively reasonable. The quantity of data justifies this. We have 5499 data points as participants opinion (13 feature models and 141 participants 3 (sub characteristics) and 130 metric values (13 feature models 10 structural metrics). As value of the independent variable we have 130 metric values and for dependent variables we obtained 5499 subjective opinions. In the following section, on the basis of collected data, we brief about the types of statistical analysis that we applied in order to better understand the existing relationships and to actually test our hypotheses. In the first step, in order to decide whether our sample comes from a population with a specific distribution or not we applied the Kolmogorov-Smirnov test [18]. As a result we realized that the data did not have a normal distribution. Thus in the second step we applied the Spearman's rho correlation. We applied this test from the following perspectives:

a) To study inter metric correlation.
b) To study the inter quality correlation.
c) Lastly to study metric indicativeness.

.

TABLE 6: INTER METRIC STUDY

Correlations												
			NF	NTOP	NLEAF	CC	CTC	ROV	COC	FOC	NVC	DT
Spearman's Rho	NF	Correlation Coefficient		.501	.941**	.152	.571*	.047	1.000**	.497	.676*	.778**
		Significance		.081	.000	.621	.042	.878	.	.084	.011	.002
	NTOP	Correlation Coefficient			.502	.208	.326	.000	.501	.342	.362	.151
		Significance			.080	.496	.276	1.000	.081	.252	.224	.623
	NLEAF	Correlation Coefficient				.000	.674*	.294	.941**	.414	.810**	.621*
		Significance				1.000	.012	.329	.000	.159	.001	.024
	CC	Correlation Coefficient					.118	-.333	.152	.182	-.128	.130
		Significance					.701	.267	.621	.551	.676	.671
	CTC	Correlation Coefficient						.320	.571*	.429	.894**	.425
		Significance						.286	.042	.144	.000	.148
	ROV	Correlation Coefficient							.047	.102	.380	-.215
		Significance							.878	.740	.200	.481
	COC	Correlation Coefficient								.497	.676*	.778**
		Significance								.084	.011	.002
	FOC	Correlation Coefficient									.281	.679*
		Significance									.352	.011
	NVC	Correlation Coefficient										.368
		Significance										.216

6. RESULTS AND ANALYSIS

A. *Inter metric correlation:*

The objective of this test was to check whether the measures were representing individual aspects or are overlapping with other metrics. The results are shown in table 6

Correlations

The values highlighted are those which are found to have a significant positive correlation. According to spearman's correlation those values are considered to meaningful where the value of significance is <.05[19]. The values highlighted in the above table are the values which have a significant degree of correlation. For example NF and NLeaf are highly correlated and also seem to be similar. We can say that if the total features in a feature model is more the number of leaf features will also be more in such feature models. Another strong correlation is found between Nleaf and COC, this means that the number of leaf features and coefficient of connectivity density is linearly correlated. In other words a highly dense feature model will have large number of leaf features, which is obvious. Similarly CTC and NVC are highly correlated which indicates that cross tree constraints of a feature model are significantly correlated to the total number of valid configurations. We can also say that a densely populated feature model will have high Number of features and leaf features.

The values of correlation obtained in this test lead us to believe that all metrics are not necessarily required to evaluate the usability of any feature model. The reason for the same is that few metrics are highly correlated and thus can also be used interchangeably.

B. *Inter quality correlation:*

This test shows us which out of the sub characteristics are correlated to each other. The results are shown in table 7

.

TABLE 7: SPEARMAN'S RHO CORRELATION FOR INTER QUALITY STUDY

			Understandability	Learnability
Spearman's Rho	Communicativeness	Correlation Coefficient	.737**	.737**
		Sig.(2-tailed)	.004	.004
	Understandability	Correlation Coefficient		.733**
		Sig.(2-tailed)		.004

TABLE 8: SPEARMAN'S RHO CORRELATIONS FOR METRIC INDICATIVENESS

		NF	NTOP	NLEAF	CC	CTC	ROV	COC	FOC	NVC	DT
Communicativeness	Correlation Coefficient	-.544	-.225	-.445	-.190	-.635*	.269	-.544	-.538	-.439	-.729**
	Significance	.055	.461	.128	.534	.020	.374	.055	.058	.133	.005
Understandability	Correlation Coefficient	-.468	-.011	-.336	-.062	-.383	.403	-.468	-.613*	-.287	-.800**
	Significance	.106	.970	.262	.840	.196	.172	.106	.026	.342	.001
Learnability	Correlation Coefficient	-.313	-.246	-.220	-.211	-.287	.307	-.313	-.670*	-.134	-.567*
	Significance	.297	.417	.471	.489	.341	.307	.297	.012	.663	.043

The test results show that out of the three selected sub characteristics of usability, communicativeness is said to be closely related to understandability and Learnability. We can thus infer that a feature model which is more communicative will be easy to understand and easy to learn. The degree of communicativeness of a feature model will lead to higher understandability and Learnability, eventually leading to high usability. The most usable feature model would be the one which has higher degree of communicativeness.

C. *Metric Indicativeness:*

This test is a major one and shows which metrics are good indicators of usability. The results are shown in table 8.

This test focuses to highlight significant relationship between the metrics and three sub characteristics. It can be seen that understandability and Depth of tree have a high degree of correlation. It means that the more the depth of the tree the lesser the level of understandability. Similarly Learnability and FOC i.e. flexibility of configuration are negatively correlated with learnability. It means that increase in number of configurations will lead to lesser understandability and learnability of the feature model. Broadly speaking we can conclude that metrics NF, CTC, COC are correlated to communicativeness; FOC and DT are correlated to both understandability and learnability. This proves our alternative hypothesis true that there is a correlation found between structural metrics and feature model usability. This correlation study will further lead us to assess the prediction power of these structural metrics. This correlation study has scrutinized that the metrics which are correlated and will be able to predict feature model usability. But since the p value of correlation is greater than .05; there is a weak correlation between the two variables. The major variation in results with those of the author's experiment may be due to two reasons. Firstly the chosen metrics were proposed by the author to assess maintainability of feature models whereas our focus of study is assessment of usability. We want to judge whether the chosen metrics are capable to assess feature model usability. Secondly, in our experiment we have chosen different feature models from that of Bagheri et.al. In both the experiments the feature models vary in size and structure.

7. DISCUSSIONS

During experimental studies various issues exist which threaten the validity of the study. Few threats to conclusion, construct, internal and external validity that exist and were taken care of during our experiment are:

Threats to conclusion validity:

This validity defines that conclusions drawn are statistically valid. In our case, sample size is an issue that could have affected this validity. Our sample size is limited because we had restricted access to participants who had appropriate knowledge of software product line domain. Also we had limited access to feature models. Although the SPLOT repository had a large number of feature models but out of them only a limited number fitted in our experimental requirement. But as the experiment is in its first stage, we will try to increase the sample size in the future stages and overcome this threat to our experimental validity.

Threats to Construct Validity:

Construct validity is the degree to which the variables are accurately measured during the experiment using correct measurement instruments. The independent variables were previously validated by Bagheri et.al in their experimental study. In our case, the dependent variable is measured on 7 point likert scale providing best number of options to the participants. We have also applied valid reliability test to check the reliability of the same.

Internal Validity:

The internal validity is the degree of confidence about the cause-effect relationship i.e. what are the factors of interest and what results have been obtained. An internally invalid experiment will lead to irrelevant results from the point of view of a causal relationship. In our case the analysis is correlation based. Also there was no difference between the subjects i.e. they all were from the computer science discipline. Feature models were selected keeping in mind various domains of the real life. The participants were provided enough time to understand and become familiar with the task. The time duration of the experiment was also short. Plagiarism was also taken care of.

External Validity:

This validity is the extent up to what level the results can be generalized to the population under study and also other scenarios of real life. In our case we tried to include feature models which were best fitting in size and covering a wide domain. We had difficulty in including professionals in our study but have included students who are from computer science domain only. However in our further experiments we will involve professionals and educationist to strengthen this validity.

8. CONCLUDING REMARKS AND FUTURE WORKS

In software product lines paradigm, feature models are often needed to depict the existing commonality and variability. Due to this reason, the quality of feature models has much importance. Literature review revealed that a set of metrics have been proposed by Bagheri et.al to assess feature model maintainability.

In this paper we have empirically analyzed these metrics to assess their correlation with feature model usability. In the process of experimentation we have employed Spearman's correlation. Results show that few metrics are correlated with the sub characteristics of usability. Therefore our alternative hypothesis that there exists a significant correlation between the structural metrics and the usability of SPL feature models is accepted. The results revealed that out of the given set of metrics; NF, CTC, and COC are correlated to communicativeness; whereas FOC and DT are correlated to both understandability and Learnability. This should be noted that correlation study is not sufficient enough to make predictions. It is only a measure which gives a probability that study can be conducted in a particular direction. It also identifies items which can be considered for further analysis. Prediction models are always needed for any kind of predictive study. As in our case we also need to apply prediction models that will identify metrics influencing feature model usability. To get more precise results we plan to propose measures keeping in mind assessment of usability.

The results confirm that although the metrics are correlated they will not predict feature model usability. A family of experiments is needed to come to firm conclusions. We are currently planning to build prediction models to state the exact degree of usability. Also a metric keeping in mind usability is under theoretical validation. Like all other empirical study, more experimentation is needed to validate our results and draw final comparison and remarks.

REFERENCES:

[1] Pohl, K., Böckle, G., & van der Linden, F. J. (2005). Software product line engineering: foundations, principles and techniques. Springer Science & Business Media.

[2] Bass, L., Clements, P., Cohen, S., Northrop, L., & Withey, J. (1997). Product Line Practice Workshop Report (No. CMU/SEI-97-TR-003). CARNEGIE-MELLON UNIV PITTSBURGH PA SOFTWARE ENGINEERING INST.

[3] Kang, K. C., Cohen, S. G., Hess, J. A., Novak, W. E., & Peterson, A. S. (1990). Feature-oriented domain analysis (FODA) feasibility study (No. CMU/SEI-90-TR-21). Carnegie-Mellon Univ Pittsburgh Pa Software Engineering Inst.

[4] Iso, I. S. O. (2001). IEC 9126-1: Software Engineering-Product Quality-Part 1: Quality Model. Geneva, Switzerland: International Organization for Standardization.

[5] Sharma, A., & Vyas, G. (2013). New Set of Metrics for Accessing Usability in Feature Oriented Programming. International Journal of Computer Applications, 81(11), 19-22.

[6] Bagheri, E., & Gasevic, D. (2011). Assessing the maintainability of software product line feature models using structural metrics. Software Quality Journal, 19(3), 579-612.

[7] Mccall, J. A. &, Richards, P. K. and Walters, G. F. (1977). Factors in Software Quality, Volumes I, II, and III. US Rome Air Development Center Reports, US Department of Commerce, USA.

[8] Boehm, B. W., Brown, H., Lipow,M. (1978) "Quantitative Evaluation of Software Quality," TRW Systems and Energy Group, 1978.N. Fenton, "Software measurement: A necessary scientific basis," IEEE Transactions on software engineering, pp. 199–206, 1994.

[9] Brito e Abreu F. and Carapuca R. (1994). Object – Oriented Software Engineering: Measuring and controlling the development process. 4th Int Conference on Software Quality.

[10] Chidamber S. and Kemerer C. (1994). A Metrics Suite for Object Oriented Design. IEEE Transactions on Software Engineering 20(6), 476 - 493.

[11] Fenton N. and Neil M. (2000). Software Metrics: a Roadmap. Future of Software Engineering. Anthony Finkelstein Ed., ACM, 359 - 370.

[12] Genero M. (2002). Defining and Validating Metrics for Conceptual Models. Ph.D. thesis, University of Castilla - La Mancha.

[13] Lorenz M. and Kidd J. (1994). Object - Oriented Software Metrics: A Practical Guide. Prentice Hall, Englewood Cliffs, New Jersey.

[14] Marchesi M. (1998). OOA Metrics for the UnifiedModeling Language. Proceedings of the 2nd Euromicrof Conference on Software Maintenance and Reengineering,67 - 73.

[15] Basili, V. R. (1992). Software modeling and measurement: the Goal/Question/Metric paradigm.

[16] Lozano, L. M., García-Cueto, E., & Muñiz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. Methodology, 4(2), 73-79.

[17] Eisinga, R.; Te Grotenhuis, M.; Pelzer, B. (2013). "The reliability of a two-item scale: Pearson, Cronbach or Spearman-Brown?". International Journal of Public Health 58 (4): 637–642.

[18] Corder, G. W.; Foreman, D. I. (2014). Nonparametric Statistics: A Step-by-Step Approach. Wiley. ISBN 978- 1118840313.

[19] Bonett DG, Wright, TA (2000). "Sample size requirements for Pearson, Kendall, and Spearman correlations". Psychometrika. 65: 23–28.