Empirical Validation of Product-line Architecture Extensibility Metrics

Edson Oliveira Jr. and Itana M. S. Gimenes
Informatics Department, State University of Maringá, Av. Colombo, 5790, Maringá PR, Brazil

Keywords: Empirical Validation, Extensibility, Metrics, Software Product Line Architecture, Variability Management.

Abstract: The software product line (PL) approach has been applied as a successful software reuse technique for specific domains. The SPL architecture (PLA) is one of the most important SPL core assets as it is the abstraction of the products that can be generated, and it represents similarities and variabilities of a PL. Its quality attributes analysis and evaluation can serve as a basis for analyzing the managerial and economical values of a PL. This analysis can be quantitatively supported by metrics. Thus, we proposed metrics for the PLA extensibility quality attribute. This paper is concerned with the empirical validation of such metrics. As a result of the experimental work we can provide evidence that the proposed metrics serve as relevant indicators of extensibility of PLA by presenting a correlation analysis.

1 INTRODUCTION

The software product line (PL) engineering approach has gained increasing attention over the last years due to competitiveness in the software development segment. The economic considerations of software companies, such as cost and time to market, motivate the transition from single-product development to the PL approach, in which products are developed in a large-scale reuse perspective (Linden et al., 2007; Capilla et al., 2013).

One of the most important assets of a PL is its architecture (PLA). The PLA plays a central role at the development of products from a PL as it is the abstraction of the products that can be generated, and it represents similarities and variabilities of a PL. PLAs provide a general notion of potential PL specific products by means of the reuse of the PL core assets. In order to derive specific products according to the company's main goals, PLAs must be evaluated. Such an evaluation may occur by taking into consideration metrics (Pohl et al., 2005; Capilla et al., 2013), which may both evidence the quality of a PL and serve as a basis to analyze the managerial and economical value of a PL (Bckle et al., 2004). As a PLA must be encompass similarities and variabilities, metrics are applied to a set of assets from which variants can be generated rather than one specific product. Therefore, specific PLA quality attributes metrics must be defined and validated to provide effective indicators with regard to the overall PL development and evolution.

In this paper it is proposed PLA metrics for the extensibility quality attribute. Extensibility is measured based on the relation between abstract classes and methods over concrete classes and methods (Sane and Birchenough, 1999; Smith, 2012). Both theoretical and empirical validations (Briand et al., 1995; Bertoa et al., 2006; García et al., 2009) are necessary to validate a set of metrics. Theoretical validation of the extensibility metrics have been done in (Oliveira Junior et al., 2008). Thus, this paper is concerned with the empirical validation of the proposed metrics for PLA extensibility quality attribute. Such a validation aims at correlating the metrics with subject's extensibility rating, respectively, when generating PLA configurations. A PLA configuration represents a specific PL product with variabilities resolved.

This remainder of this paper is organized as follows: Section 2 presents the extensibility metrics to be validated; Section 3 presents how the experimental study was planned and carried out to validate the metrics; Section 4 discusses the results obtained in this study; and Section 5 provides the conclusions and directions for future work.

2 EXTENSIBILITY METRICS FOR SOFTWARE PRODUCT LINE ARCHITECTURES

Understand extensibility is essential from the PL

Oliveira Jr. E. and M. S. Gimenes I..
Empirical Validation of Product-line Architecture Extensibility Metrics.
DOI: 10.5220/0004745201110118
In *Proceedings of the 16th International Conference on Enterprise Information Systems* (ICEIS-2014), pages 111-118
ISBN: 978-989-758-028-4

adoption point as a PL manager is able to analyze the extensibility of the potential PL specific products to be produced.

Organizations which have a developed PL core asset for a certain domain can analyze the extensibility of the distinct configurations and the PL evolution. Therefore, a PL manager may choose from a set of feasible configurations which are the most interesting to be produced.

The extensibility metrics for PLA were composed based on the following extensibility principles (Sane and Birchenough, 1999; Woolf, 1997; Smith, 2012): the number of abstract methods divided by the total number of methods (concrete plus abstract) of a class. Each metric measures the extensibility of class, interface and component based on one of the following PL variability concepts:

- **Variability**, according to Capilla et al. (Capilla et al., 2013), is "the ability of a software or artifact to be changed, customized or configured for use in a particular context." Although a variability can take place at different levels of abstraction and artifacts, the extensibility metrics in this paper address only class and component UML artifacts that result from PL activities (Oliveira Junior et al., 2013; Oliveira Junior et al., 2010) and represents the PLA;
- **Variation Point** is the resolution of variabilities in generic artifacts of a PL. According to Capilla et al., (Capilla et al., 2013), "a variation point identifies one or more locations at which the variation will occur." Thus, a variation point may take place at generic artifacts and at different levels of abstraction. Basically, a variation point answers the question: What varies in a PL? (Pohl et al., 2005; Capilla et al., 2013; Linden et al., 2007); and
- **Variant** represents the possible elements through which a variation point may be resolved. It may also represent a way to directly resolve a variability. Basically, a variant answers the question: How does a variability or a variation point vary in a PL? (Pohl et al., 2005; Capilla et al., 2013).

Figure 1 is an excerpt of a sorting feature, which has a variability aimed at sorting elements by selecting a proper algorithm. The variation points (`SortingElement` and `SortingAlgorithm`) are annotated with the stereotype ≪variationPoint≫. The variants are annotated with the stereotypes ≪alternative_OR≫ (`NumberElement`, `StringElement`, `QuickSort`, `HeapSort` and `BubbleSort`) and ≪mandatory≫ (`MainSortProgram`). Detailed information with regard to the stereotypes used to annotate variability concepts can be found in (Oliveira Junior et al., 2010; Oliveira Junior et al., 2013). All classes and interfaces from the `algorithms` package form the component `sorting`.

The extensibility metrics taken into consideration in this paper are as follows:

***ExtensInterface*:** measures the extensibility of an interface. It always has value 1.0 as an interface has only abstract methods. It means that interfaces are 100% extensible by means of their abstract methods. This metric is represented by the following formula (1):

$$\left.\begin{array}{l}\textbf{ExtensInterface(Itf)} = nAbs/(nConc + nAbs) = nAbs/(0 + nAbs) = nAbs/nAbs = 1.0\\ \text{where:}\\ \bullet\ nAbs = \#\ of\ abstract\ methods\ of\ an\ interface\ (Itf)\\ \bullet\ nConc = \#\ of\ concrete\ methods\ of\ an\ interface\ (Itf)\end{array}\right\} \quad (1)$$

***ExtensClass*:** measures the extensibility of a class. This metric is represented by the following formula:

$$\left.\begin{array}{l}\textbf{ExtensClass(Cls)} = nAbs/(nConc + nAbs)\\ \text{where:}\\ \bullet\ nAbs = \#\ of\ abstract\ methods\ of\ a\ class\ (Cls)\\ \bullet\ nConc = \#\ of\ concrete\ methods\ of\ a\ class\ (Cls)\end{array}\right\} \quad (2)$$

***ExtensVP*:** measures the extensibility of a variation point. It is the value of the metric *ExtensClass* (Equation 2) for a class which is a variation point or the value of the metric *ExtensInterface* (Equation 1) for an interface which is a variation point, plus the sum of the *ExtensClass* (Equation 2) value for each associated variant class. This metric is represented by the following formula:

$$\left.\begin{array}{l}\textbf{ExtensVP(X)} = \begin{cases} ExtensClass(X) + \sum_{i=1}^{n} ExtensClass(Ass_i) \text{ if X is a class}\\ ExtensInterface(X) + \sum_{i=1}^{n} ExtensClass(Ass_i) \text{ if X is an interface}\end{cases}\\ \text{where:}\\ \bullet\ n = \#\ of\ (inclusive + exclusive + optional + mandatory)\ variant\ classes\ and\ interfaces\ associated\ (Ass)\end{array}\right\} \quad (3)$$

***ExtensVariability*:** measures the extensibility of a variability. It is the sum of the metric *ExtensVP* (Equation 3), for each variation point. This metric is represented by the following formula:

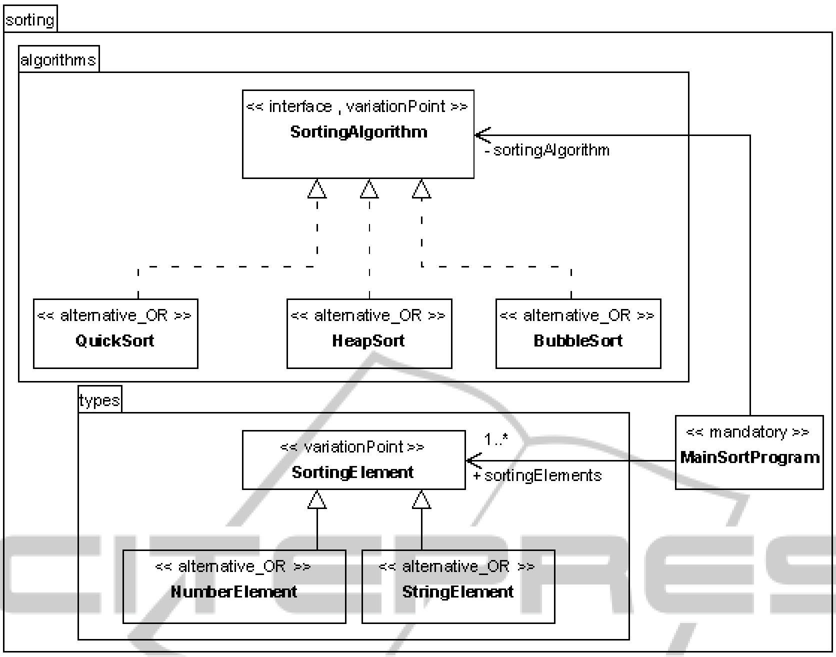

Figure 1: Excerpt of a Sorting Feature.

$$\textbf{ExtensVariability(Vbt)} = \sum_{i=1}^{nVP} ExtensVP(Cls_i) \quad (4)$$

where:

- *nVP = # of class and interface (Cls) variation points*

ExtensComponent**:** measures the extensibility of a variable PLA component. It is the sum of the metric *ExtensVariability* (Equation 4), for each variability in a component. This metric is represented by the following formula:

$$\textbf{ExtensComponent(Cpt)} = \sum_{i=1}^{nVar} ExtensVariability(Var_i) \quad (5)$$

where:

- *nVar = # of variabilities (Var) in a component (Cpt)*

ExtensPLA**:** measures the extensibility of a PLA. It is the sum of the *ExtensComponent* (Equation 5) for each component of a PLA. This metric is represented by the following formula:

$$\textbf{ExtensPLA(PLA)} = \sum_{i=1}^{nCpt} ExtensComponent(Cpt_i) \quad (6)$$

where:

- *nCpt = # of PLA variable components*
- *Cpt_i is the i_{th} component of a PLA*

3 EXPERIMENTAL STUDY

In this section we describe the experiment carried out to empirically validate the proposed metrics as indicators of PLA extensibility (Basili et al., 2007).

3.1 Definition

The goal of the experiment is presented as follows:

Analyze *collected metrics from UML models*
For the purpose of *validating*
With respect to *the capability to be used as PLA extensibility indicators*
From the point of view of *software product line architects*
In the context of *graduate students of the Software Engineering area at the University of Waterloo (UWaterloo), University of So Paulo (ICMC-USP), and State University of Maring (UEM).*

3.2 Planning

3.2.1 Context Selection

The experiment was carried out in an academic environment.

3.2.2 Selection of Subjects

A group of Software Engineering graduate students from ICMC-USP, UEM, and UWaterloo. They have experience in the design of product lines and variabilities using UML.

3.2.3 Variable Selection

The independent variables were:

- **extensibility**, which is a factor with two treatments
  - the extensibility metrics; and
  - the extensibility subject's rating.
- **PLA**, which is a pre-fixed variable with value "Arcade Game Maker (AGM)".

The dependent variable was the **extensibility correlation** between the extensibility metrics and the subject's extensibility rating provided by each subject.

3.2.4 Instrumentation

The following objects compose the instrumentation:

- a document describing the Arcade Game Maker (AGM) PL (SEI, 2010);
- the AGM UML class and component models;
- an AGM traceability model from classes to components; and
- a resolution model containing the AGM variabilities to be resolved at class level.

3.2.5 Hypotheses Formulation

The following hypotheses were defined to be tested in this study:

- **Null Hypothesis (H_0):** There is no significant correlation between the PLA extensibility metric (*ExtensPLA*) and the subject's extensibility rating (*ExtensSubjectRate*) for some PLA configuration:

$$H_0 : \mu_{(ExtensPLA)} \neq \mu_{(ExtensSubjectRate)}$$

- **Alternative Hypothesis (H_1):** There is a significant correlation between the PLA extensibility metric (*ExtensPLA*) and the subject's extensibility rating (*ExtensSubjectRate*) for some PLA configuration:

$$H_1 : \mu_{(ExtensPLA)} = \mu_{(ExtensSubjectRate)}$$

3.2.6 Experiment Design

This experiment design defined was: one factor with two treatments by performing a correlation analysis between such treatments (Wohlin et al., 2010).

3.3 Operation

3.3.1 Preparation

When the experiment was carried out, all of the subjects had graduated in the Software Engineering area, in which they have learned how to design at least object-oriented (OO) class diagrams using UML. In addition, all of the subjects had experience in applying PL and variability concepts to OO systems designed using UML.

The material prepared to the subjects consisted of:

- the class diagram representing the core asset of the AGM PL;
- the AGM component diagram, representing its logical architecture;
- an AGM traceability model from classes to components;
- the description of the AGM PL;
- the SMartyProfile (Oliveira Junior et al., 2010), which is a UML metamodel, thus the subjects can understand how the variabilities are represented in class and component diagrams;
- a variability resolution model, which the subjects could resolve the variabilities to generate one AGM configuration; and
- a test (questionnaire) describing extensibility concepts, which the subjects had to rate the associated extensibility of each generated AGM configuration based on linguistic labels (Table 1).

Table 1: Linguistic Labels for Subject's Extensibility Rating.

Extremely Low	Low	Neither Low nor High	High	Extremely High

We selected five linguistic labels, based on Bonissone (Bonissone, 1980), as we considered they are significant to cover the extensibility category of our variables and bring out balance to obtain better results.

3.3.2 Execution

The subjects were given the material described in *Preparation* (Section 3.3.1). It was required to each subject to generate one AGM configuration. It was done by following instructions on how to resolve the AGM variability resolution model, and how to rate the extensibility associated to the configurations generated from the subjects view point. All the tasks were performed by each subject alone, with no time limit to solve them and neither sequentially nor simultaneously.

As the metric ExtensPLA is a composition of the remaining extensibility metrics of this paper, we only take ExtensPLA into consideration for the validation purpose. In addition, the ExtensPLA value of each configuration was divided by the ExtensPLA value of the overall AGM PLA, thus resulting in a value ranging from 0.0 to 1.0.

3.3.3 Data Validation

The tasks performed by the subjects were collected.

We consider the subjects subjective evaluation reliable based on their characterization.

3.4 Analysis and Interpretation

We summarized the collected data by calculating the metrics ExtensPLA for the thirty AGM configurations generated by the subjects, as well as verifying the extensibility rating of such configurations. Table 2 presents the observed values for the ExtensPLA metric from the generated AGM configurations.

Table 2: Observed Values for the ExtensPLA Metric from the Generated Configurations.

Config. #	ExtensPLA	Config. #	ExtensPLA
1	0.81	16	1.00
2	0.61	17	0.80
3	1.00	18	0.61
4	0.80	19	0.61
5	0.80	20	0.61
6	0.61	21	0.80
7	0.80	22	0.80
8	0.61	23	0.61
9	1.00	24	0.80
10	1.00	25	1.00
11	0.81	26	0.61
12	1.00	27	0.61
13	0.61	28	0.61
14	0.61	29	0.80
15	1.00	30	0.61

3.4.1 Descriptive Statistics

Figure 2 presents the ExtensPLA descriptive statistics for the observed values of Table 2.

3.4.2 Normality Tests

We can clearly observe that the ExtensPLA values distribution (Figure 2) is non-normal. In spite of it, Shapiro-Wilk and Kolmogorov-Smirnov normality tests were conducted to make sure of it.

The following hypotheses were proposed for both normality tests with regard to the ExtensPLA metric:

- **Null Hypothesis (H_0):** the ExtensPLA observed values distribution is normal, i.e., the significance value (p) is greater than 0.05 ($p > 0.05$); and
- **Alternative Hypothesis (H_1):** the ExtensPLA observed values distribution is non-normal, i.e., the significance value (p) is less or equal to 0.05 ($p \leq 0.05$).

Taking into account a sample size (N) of 30, with mean (μ) 0.7390, standard deviation (σ) 0.1487, and median ($\tilde{x}$) 0.7060, the *ExtensPLA* metric obtained a significance value:

- $p = 0.009$ ($0.009 < 0.05$) for the *Kolmogorov-Smirnov* test;
- $p = 0.000011$ ($0.000011 < 0.05$) for the *Shapiro-Wilk* test.

Thus, there is evidence, for both normality tests, that the null hypothesis (H_0) must be rejected at a significance level of 5%. Then, we cannot consider the ExtensPLA observed values distribution normal and, consequently, a non-parametric statistic method must be used to analyze the data.

3.4.3 Spearman's Rank Correlation

As ExtensPLA distribution is non-normal, we applied the non-parametric Spearman's Correlation (ρ) (Spearman, 1904) to support the interpretation of the data. This method allows establishing whether there is a correlation between two sets of data, in our case, for the ExtensPLA and the Subject's Extensibility Rating. Equation (7) presents the Spearman's ρ formula:

$$\left.\begin{array}{l}\rho = 1 - \frac{6}{n(n^2-1)} \sum_{i=1}^{n} d_i^2, \text{ where } n \text{ is the sample size } (N)\end{array}\right\} \quad (7)$$

In this study, it was performed the following correlation (Corr.1): **ExtensPLA and the subjects extensibility rating**, which shows that the understanding of extensibility by the subjects corroborates the ExtensPLA metric, aiming at providing evidence of an indicator for PLA extensibility.

Table 3 presents the Spearman's ranking correlation for Corr.1. The Spearman ρ coefficient (Equation

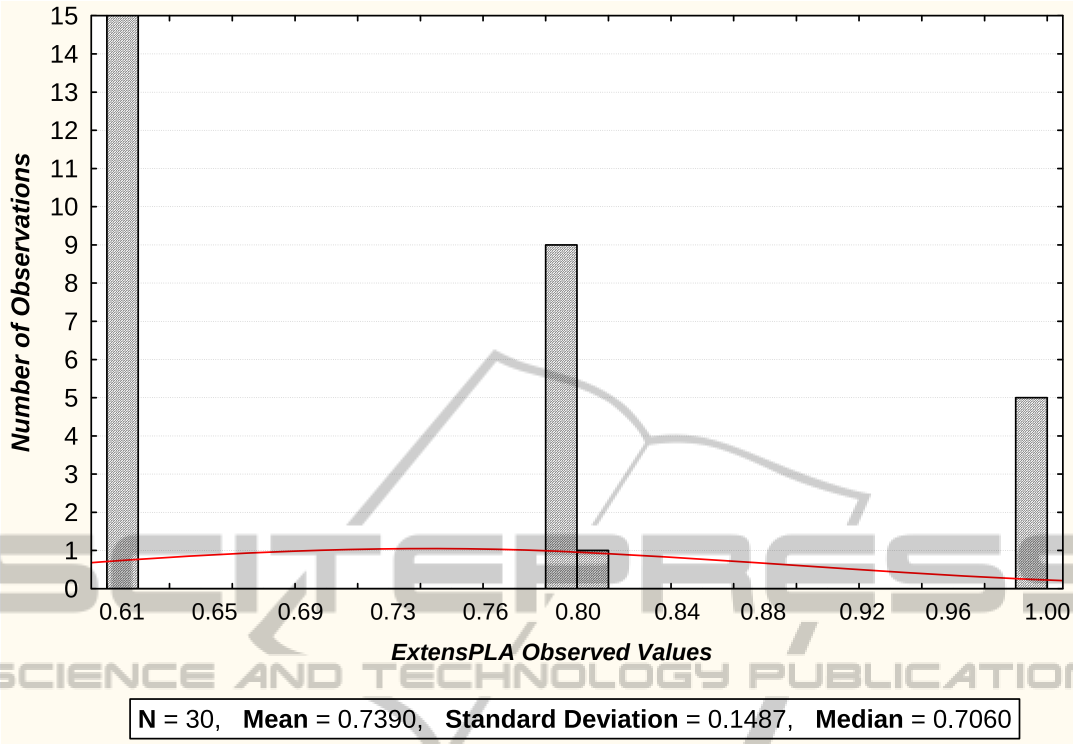

Figure 2: Descriptive Statistics for the ExtensPLA Observed Values.

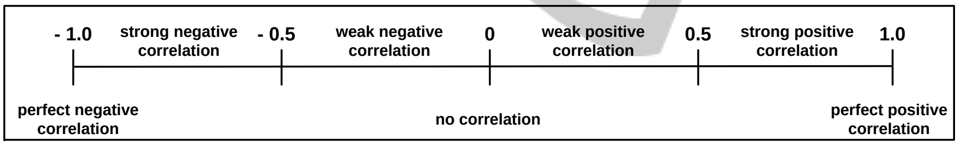

Figure 3: Spearman's Rank Correlation Scale.

7) for Corr.1 is calculated as follows:

$$\rho(_{Corr.1}) = 1 - \frac{6}{30(30^2-30)} \times 711.23 = 1 - \frac{6}{26970} \times 711.23 = 1 - 0.1582 = \mathbf{0.8413}$$

Thus, according to Figure 3, there is a strong positive correlation ($\rho(_{Corr.1})$ = 0.84) between the metric ExtensPLA and the Subject's Extensibility Rating.

Based on the proposed correlation, we have evidence to reject the null hypothesis H_0 of the study, and accept the alternative hypothesis H_1 (Section 3.2.5), which states that extensibility metrics are significantly correlated to the subject's extensibility rating.

3.5 Validity Evaluation

In this section we discuss the empirical study's threats to validity and how we tried to minimize them.

3.5.1 Threats to Conclusion Validity

The only issue that we take into account as a risk to affect the statistical validity is the sample size (N=30), which can be increased during prospective replications of this study in order to reach normality of the observed values and generalize results.

3.5.2 Threats to Construct Validity

We proposed subjective metrics for measuring the subject's extensibility rating, as linguistic labels. As the subjects have experience in modeling OO systems using at least class diagrams, we take their ratings as significant. The construct validity of the extensibility metrics used as independent variables is guaranteed by some insights carried out on a previous study of metrics for PLA (Oliveira Junior et al., 2008).

Table 3: Spearman's Correlation for Corr.1: ExtensPLA and Subjects Extensibility Rates.

Config. #	*ExtensPLA*	r_a	Subject's Extensibility Rating	r_b	d $\|r_a - r_b\|$	d^2
1	0.61	23	High	17.5	5.5	30.25
2	0.61	23	High	17.5	5.5	30.25
3	0.81	6	Extremely High	3.5	2.5	6.25
4	0.80	11	High	17.5	6.5	42.25
5	0.61	23	High	17.5	5.5	30.25
6	1.00	3	Extremely High	3.5	0.5	0.25
7	0.61	23	High	17.5	5.5	30.25
8	0.61	23	High	17.5	5.5	30.25
9	0.80	11	High	17.5	6.5	42.25
10	1.00	3	Extremely High	3.5	0.5	0.25
11	1.00	3	Extremely High	3.5	0.5	0.25
12	0.61	23	High	17.5	5.5	30.25
13	0.61	23	High	17.5	5.5	30.25
14	0.61	23	High	17.5	5.5	30.25
15	0.80	11	High	17.5	6.5	42.25
16	1.00	3	Extremely High	3.5	0.5	0.25
17	0.80	11	High	17.5	6.5	42.25
18	0.80	11	High	17.5	6.5	42.25
19	0.61	23	High	17.5	5.5	30.25
20	0.80	11	High	17.5	6.5	42.25
21	0.61	23	High	17.5	5.5	30.25
22	1.00	3	Extremely High	3.5	0.5	0.25
23	0.61	23	High	17.5	5.5	30.25
24	0.61	23	High	17.5	5.5	30.25
25	0.80	11	High	17.5	6.5	42.25
26	0.61	23	High	17.5	5.5	30.25
27	0.61	23	High	17.5	5.5	30.25
28	0.80	11	High	17.5	6.5	42.25
29	0.61	23	High	17.5	5.5	30.25
30	0.80	11	High	17.5	6.5	42.25

3.5.3 Threats to Internal Validity

We dealt with the following issues:

- **Differences Among Subjects.** As we dealt with a small sample, variations in the subject skills were reduced by applying the within-subject task design. Thus, subjects experiences had approximately the same degree with regard to UML modeling, and PL and variabilities basic concepts.
- **Accuracy of Subject Responses.** Extensibility was rated by each subject. As they have medium experience in UML modeling, and PL and variabilities concepts, we considered their responses valid.
- **Fatigue Effects.** On average the experiment lasted for 58 minutes, thus fatigue was considered not very relevant. Also, the variability resolution model contributed to reduce such effects.
- **Measuring PLA and Configurations.** As PLA can be analyzed based on its products (configurations), measuring derived configurations provide a means to analyze PLA quality attributes by allowing the performing of trade-off analysis to prioritize such attributes. Thus, we consider valid the application of the metrics to PLA configurations to rate the overall PLA extensibility.
- **Other Important Factors.** Influence among subjects could not really be controlled. Subjects took the experiment under supervision of a human observer. We believe that this issue did not affect the study validity.

3.5.4 Threats to External Validity

Based on the greater the external validity, the more the results of an empirical study can be generalized to actual software engineering practice, two threats to validity have been identified, which are:

- **Instrumentation.** We tried to use representative class and component diagrams of real cases. However, the PL used in the experiment is non-commercial, and some assumptions can be made on this issue. Thus, more empirical studies taking a "real PL" from software organizations must be done.
- **Subjects.** Obtaining well-qualified subjects was difficult, thus we used advanced students from the Software Engineering academia. More experiments with industry practitioners and professionals must be carried out allowing us to generalize the study results.

4 DISCUSSION OF RESULTS

Obtained results of the study provided evidence that the metric ExtensPLA is a relevant indicator of PLA extensibility based on its correlation to the subject's rating.

Several more experiments must be carried out, as well as more PLA configurations must be both derived and incorporated to enhance the conclusions. In addition, we need to apply our metrics to a commercial PL in order to reduce external threats to the study validity and for gathering real evidence that these metrics can be used as extensibility indicators.

Existing literature presents no work directly related to this paper. Although some theoretical validations are presented in the literature for PLA metrics, as far as we know, no empirical validation has been performed similarly to the carried out study.

5 CONCLUSION

Current literature claims the need of metrics to allow PL architects empirically analyze the potential of a PLA, as well as PL managers analyze the aggregated managerial and economical values of a PL throughout its products.

Performing empirical validation of metrics is essential to demonstrate their practical usefulness. The proposed metrics for the extensibility (ExtensPLA) PLA quality attribute were empirically validated based on their application to a set of 30 products generated by experiment subjects from the *Arcade Game Maker* (AGM) PL. The observed metric values were submitted to normality tests which proved their non-normality. Then, Spearman's rank correlation was used to demonstrate the metrics correlations, which is: ExtensPLA has a strong positive correlation with the subject's extensibility rating.

Although we have used a non-commercial PL to conduct our experiments, we had evidence that our proposed metrics can be used as relevant indicators of extensibility of a PLA based on its derived products.

We are currently proposing changes on various issues to improve our experiments with metrics, which are: (i) increase the derived configurations sample size, which is important to stay closer to real projects and to generalize the results; (ii) conduct experiments in a more controlled environment; (iii) deal with real data from commercial PL obtained from industry; and (iv) recruit more subjects from the Software Engineering area, both from academic and industrial environments.

REFERENCES

Basili, V. R., Rombach, D., Schneider, K., Kitchenham, B., Pfahl, D., and Selby, R. W. (2007). *Empirical Software Engineering Issues: Critical Assessment and Future Directions*. Springer, New York, NY, USA.

Bckle, G., Clements, P., McGregor, J. D., Muthig, D., and Schmid, K. (2004). Calculating ROI for Software Product Lines. *IEEE Software*, 21(3):23–31.

Bertoa, M. F., Vallecillo, A., and Garca, F. (2006). *An Ontology for Software Measurement*, chapter 6, pages 175–196. Springer.

Bonissone, P. P. (1980). A Fuzzy Sets Based Linguistic Approach: Theory and Applications. In *Proceedings of Conference on Winter Simulation*, pages 99–111, Piscataway, NJ, USA. IEEE Press.

Briand, L., Emam, K. E., Morasca, S., El, K., and Morasca, E. S. (1995). Theoretical and Empirical Validation of Software Product Measures. ISERN-95-03, International Software Engineering Research Network.

Capilla, R., Bosch, J., and Kang, K.-C. (2013). *Systems and Software Variability Management - Concepts, Tools and Experiences*. Springer, New York, NY, USA.

García, F., Ruiz, F., Calero, C., Bertoa, M. f., Vallecillo, A., Mora, B., and Piattini, M. (2009). Effective use of ontologies in software measurement. *Knowl. Eng. Rev.*, 24(1):23–40.

Linden, F. J. v. d., Schmid, K., and Rommes, E. (2007). *Software Product Lines in Action: The Best Industrial Practice in Product Line Engineering*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Oliveira Junior, E. A., Gimenes, I. M. S., and Maldonado, J. C. (2008). A Metric Suite to Support Software Product Line Architecture Evaluation. In *Proceedings of the Conferencia Latinoamericana de Informtica*, pages 489–498, Santa F, Argentina.

Oliveira Junior, E. A., Gimenes, I. M. S., and Maldonado, J. C. (2010). Systematic Management of Variability in UML-based Software Poduct Lines. *Journal of Universal Computer Science (J.UCS)*, 16(17):2374–2393.

Oliveira Junior, E. A., Gimenes, I. M. S., Maldonado, J. C., Masiero, P. C., and Barroca, L. (2013). Systematic Evaluation of Software Poduct Line Architectures. *Journal of Universal Computer Science (J.UCS)*, 19(1):25–52.

Pohl, K., Böckle, G., and Linden, F. J. v. d. (2005). *Software Product Line Engineering: Foundations, Principles, and Techniques*. Springer-Verlag, Secaucus, NJ, USA.

Sane, A. and Birchenough, A. (1999). First Class Extensibility for UML - Packaging of Profiles, Stereotypes, Patterns. In *Proceedings of the International Conference on the Unified Modeling Language*, pages 265–277. Springer-Verlag.

SEI (2010). Arcade Game Maker Pedagogical Product Line.

Smith, J. M. (2012). *Elemental Design Patterns*. Addison-Wesley Professional, New York, NY, USA.

Spearman, C. (1904). The Proof and Measurement of Association Between Two Things. *American Journal of Psychology*, 15(1):72–101.

Wohlin, C., Runeson, P., Hust, M., Ohlsson, M. C., Regnell, B., and Wesslun, A. (2010). *Experimentation in Software Engineering: an Introduction*. Kluwer Academic Publishers, Norwell, MA, USA.

Woolf, B. (1997). The Abstract Class Pattern. In *Proceedings of the Pattern Languages of Programming Conference*, pages 1–8. ACM.