# Size, Coupling and Cohesion Metrics for Product-Line Architecture Evaluation: Proposal and Experimental Validation

### Andre F.R. Cordeiro and Edson OliveiraJr

Department of Informatics, State University of Maringa (UEM) Maringa, Parana, Brazil

Article history Received: 23-12-2017 Revised: 14-02-2018 Accepted: 22-03-2018

Corresponding Author: Edson OliveiraJr Department of Informatics, State University of Maringa (UEM) Maringa, Parana, Brazil Email: edson@din.uem.br

Abstract: Software Product Line (SPL) is an approach for de facto reusing of software artifacts for a given domain. Amongst the artifacts developed in the context of SPL is the Product-Line Architecture (PLA). It is one of the central artifacts of the SPL core assets responsible for abstracting a common architecture for specific products. Organizations should continuously evaluate the quality of their products by managing their PLA evolution and variabilities. Thus, the PLA evaluation should be taken into consideration as one of the most important activities throughout a SPL life cycle. Existing literature presents different PLA evaluation methods, from which metrics are most used. Thus, metrics allow different PLA quality attributes to be prioritized when deriving existing products, most of the time by trading-off them. Size, Coupling and Cohesion are examples of important attributes related to Maintainability from ISO/IEC 25010. Metrics for size, coupling and cohesion have successfully been applied in SPL, however, not taking into account commonalities and variabilities. Therefore, this paper proposes metrics for measuring Size, Coupling and Cohesion of PLA, as well as their experimental validations. Such validations are performed correlating the values collected for each metric to participants rating for each quality attribute. Results indicate a weak positive correlation for the size and coupling metrics and a weak negative correlation for the Cohesion metric. In view of the presented evidence, new studies and/or investigations must be carried out to make results generalizable.

**Keywords:** Software Product Line, Product-Line Architecture, Product-Line Architecture Metrics, Experimental Validation

#### Introduction

Software Product Line (SPL) is a systematic and planned reuse approach of software artifacts in a specific domain of activity (Capilla *et al.*, 2013). The benefts of the SPL adoption are observed in a medium term (Bosch *et al.*, 2015).

In the development, maintenance and evolution of an SPL, many artifacts can be created. Amongst these artifacts is the Product-Line Architecture (PLA). The PLA represents the abstraction of the possibly instantiations of SPL products.

This characteristic makes the PLA to play a central role in the development of such products (Bass *et al.*, 2012). Considering the inuence of the PLA for the SPL approach, it is important to such an artifact be evaluated in a systematic manner. The literature presents different

methods for the evaluation of PLAs (Olumofin and Misic, 2005; OliveiraJr *et al.*, 2013).

Software metrics can support PLA evaluations (OliveiraJr *et al.*, 2008; 2013). The use of software metrics helps in the understanding, controlling and improvement of activities and/or artifacts (Fenton and Bieman, 2014). In the context of PLA, the literature presents an extensive number of metrics (Montagud *et al.*, 2012; OliveiraJr *et al.*, 2008).

In spite of the extensive number of PLA evaluation methods presented in the literature, improvement possibilities are still observed. Most of such methods in the context of SPL and PLA do not evaluate specifc attributes and/or characteristics of an SPL, such as commonalities and variabilities. Usually, such methods evaluate other attributes, such as Size, Coupling and Cohesion (Olumofin and Misic, 2005; Kim *et al.*, 2008;



© 2018 Andre F. R. Cordeiro and Edson OliveiraJr. This open access article is distributed under a Creative Commons Attribution (CC-BY) 3.0 license. Colanzi *et al.*, 2014; Colanzi and Vergilio, 2014; Oizumi *et al.*, 2012), not taking into consideration explicit variabilities. Metrics for these attributes related to maintainability of single-products are a major portion of literature (Montagud *et al.*, 2012).

Even with the evidence of the importance of such metrics for the evaluation of software artifacts, SPL/PLA context should consider commonalities and/or variabilities. In this context, the adaptation of traditional software metrics to the context of SPL can be performed (OliveiraJr and Gimenes, 2014; Marcolino *et al.*, 2013a; OliveiraJr *et al.*, 2010).

This paper proposes the adaptation of existing singleproduct metrics and their experimental validation for PLA maintainability by means of reusability and modularity attributes according to ISO/IEC 25010 (2011) (see Figure 1). Such metrics can be used to evaluate important attributes for SPL development taking into account commonalities and variabilities of a PLA. Note that this paper does not concern on how to perform quality attributes prioritization trade-o analysis based on metrics. For this purpose, please refer to OliveiraJr *et al.* (2013).

# Background

This section discusses the essential concepts of variability management and SPL metrics, which take into consideration variability.

# Variability Management

The core assets include artifacts with associated variability. Variability is the ability to customize an artifact, allowing it to be used in different instantiated products of the same SPL. The representation of the variabilities can be made by variation points and variants. Variation points are the resolution locations of a given variability in a given artifact (Capilla *et al.*, 2013). Variants allow the resolution of a given variability (Galster *et al.*, 2014). By means of the selected variants, it is possible to observe the differences between the possible products instantiated in an SPL.

Considering the characteristics of variability, it is important that they be managed in a systematic way. The literature presents different methods and/or approaches for Variability Management (VM) (Galster *et al.*, 2014; Thurimella and Bruegge, 2012; Chen *et al.*, 2009). Amongst the existing methods and/or approaches, the Stereotype-based Management of Variability (SMarty) approach (OliveiraJr *et al.*, 2010; Marcolino *et al.*, 2017; Bera *et al.*, 2015; Marcolino *et al.*, 2013b) stands for enabling representation of variability in Unified Modeling Language (UML) diagrams.

The SMarty approach supports use case, class, sequence, component and activity diagrams. In these

diagrams, one can represent and manage explicit variabilities, as well as variation points and associated variants. For the representation of these elements, an UML profle, named SMartyProfle, has been established. This profile defines stereotypes that can be applied to the elements of UML diagrams supported by SMarty.

For managing variabilities, variation points and variants, a set of guidelines, named SMartyProcess, has also been established. The guidelines help in the identification and representation of variabilities, as well as to identify the variability implementation mechanisms and analysis of product configurations.

The proposed metrics presented in this study can be applied to PLAs modeled according to SMarty, more specifically class and component diagrams representing a PLA.

# Software Product Line Metrics

The literature presents an extensive number of measures and software metrics used in the context of SPL. In Montagud *et al.* (2012), a systemic review on quality attributes and measures for SPL is presented. In total, 165 measures related to 97 different quality attributes are presented.

The number of measures and metrics observed in the work of Montagud *et al.* (2012) evidenced the extensive amount of measurements and methods used in the context of SPL. In addition, to the extended number, other characteristics are also observed. For example, 92% of the measures evaluates quality attributes related to Maintenance. Figure 2 presents ISO/IEC 25010 quality attributes for single-product quality. We can observe Maintainability represented by Modularity, Reusability, Analyzability, Modifiability and Testability.

In addition to the characteristics observed in (Montagud *et al.*, 2012), it is also possible to observe the context of the origin of measures and metrics used in SPL. Some of them have been defined for other contexts and/or paradigms, such as Object-Orientation (OO) and are used in SPL (Colanzi and Vergilio, 2014; Colanzi *et al.*, 2014; Oizumi *et al.*, 2012; Ribeiro *et al.*, 2010). The use of measures in SPL, from other contexts, does not characterize a problem as they target original attributes.

More general software product measures and metrics originally defined in other contexts/paradigms can support deriving specific metrics and measures for SPL. Such measures and metrics can be adapted to the context of SPL, allowing the evaluation of commonalities, variabilities, variation points and variants of PLAs, for example. Thus, adapted metrics could assess quality attributes of PLAs.

To do this, next section presents important concepts on single-product size, coupling and cohesion related to Modularity and Reusability attributes, then Section 3 presents our metrics proposal.



Fig. 1: ISO/IEC 25010 quality attributes (ISO/IEC,25010, 2011)



**Fig. 2:** MM1: Excerpt of the mobile media PLA

#### Size, Coupling and Cohesion Attributes

The Size attribute evaluates how much of a particular entity exists in an element/artifact (Fenton and Bieman, 2014). In the context of OO, more specifically on classes, we can consider the number of attributes and methods of a class (SFPD, 2017). Such element/artifact can be used at different stages of the software development process, for instance, design and implementation artifacts. According to Chidamber and Kemerer (1994), classes with a large number of methods tend to be more specific, limiting the possibility of reuse. Therefore, we can understand that Size can be used to estimate the Reusability (ISO/IEC,25010, 2011) of an artifact. In PLA, such attribute may support identifying larger architectural elements, which may present greater diffculties in reusing. In addition, such architectural elements may evidence other architectural problems, such as high coupling and low cohesion.

The Coupling attribute is defined as the level of interdependence among distinct software modules. It is the level in which units (different classes, packages, components) are connected (SFPD, 2017). Different coupling types are presented by ISO terminology (ISO/IEEE/IEC, 2010).

Another attribute considered in this study is Cohesion. It is usually associated with Coupling. Cohesion is defined as the number of internal elements of a design unit that are logically related to each other (SDMetrics, 2017). Different types of cohesion are presented in (ISO/IEEE/IEC, 2010).

Both Coupling and Cohesion can be verified in different elements/artifacts, from models (classes, interfaces, packages and components), to source code. Considering Coupling and Cohesion characteristics, they can be used to estimate important attributes, such as Modularity (ISO/IEC,25010, 2011). A high coupling level is detrimental to modularity and prevents a module reuse (Chidamber and Kemerer, 1994).

In PLAs Size, Coupling and Cohesion attributes can support, respectively, the evaluation of quantity, interdependence and logical relationships among architectural elements, enabling corrective and/or adaptive activities to be performed. Such activities may directly inuence PLA maintenance and evolution.

# Size, Coupling and Cohesion Metrics for PLA

This section presents metrics for Size, Coupling and Cohesion attributes. These metrics were adapted from other metrics presented in the literature, which are used in the context of SPL.

# Metric for PLA Size

This metric counts the number of operations in all PLA classes or interfaces with variability. Classes/interfaces

with no variability associated are not counted as they are present in every derived product. In addition, set and get operations, as well as overwritten operations and constructors are not counted. To do so, the metric *Number of Operations* (NumOps) (Colanzi *et al.*, 2014; Colanzi and Vergilio, 2014) is taken into account.

Equation 1 presents the Size metric NSOPLA:

$$NSOPLA = \sum_{i=1}^{n} NSO$$
 (1)

Where:

*n* = Number of PLA classes/interfaces *NSO* = Number of PLA class/interface operations

Figure 3 presents an excerpt of the Mobile Media PLA, named MM1. In such figure we can observe three optional classes: Copyist with one operation counted, Favourite with five operation counted and Sorter with one operation counted. Thus, NSOPLA (MM1) = 7. Remaining classes are mandatory and, therefore, are not taken into consideration for NSOPLA.

# Metric for PLA Coupling

This metric counts the total DepIn and DepOut (SDMetrics, 2017; Osman *et al.*, 2013) of a PLA by means of its components. DepIn is the number of dependencies of a supplier component. DepOut is the number of dependencies of a client component. Equation 2 presents the Coupling metric DepCompVariabilityPLA:

$$DepCompVariabilityPLA = \sum_{i=1}^{n} Dep \operatorname{I} n + \sum_{i=1}^{n} DepOut$$
(2)

Where:

n = Number of PLA components with variabilites DepIn = Number of dependencies of a supplier

DepOut = Number of dependencies of a client component

Figure 3 presents MM2, an excerpt of the Mobile Media PLA components. Component FavouriteMgr is optional, whereas MediaMgr is an inclusive variation point with variants MusicMgr and VideoMgr. Remaining components are mandatory:

DepIn of FavouriteMgr is 1, whereas DepOut is 0. MediaMgr has DepIn 2 and DepOut 2. MusicMgr has DepIn 1 and DepOut 0. VideoMgr has DepIn 1 and DepOut 0. Thus DepCompVariabilityPLA(MM2) = (DepIn(FavouriteMgr) DepIn(MediaMgr) ++DepIn(MusicMgr) + DepIn(VideoMgr)) +(DepOut(FavouriteMgr) +DepOut(MediaMgr) +DepOut(MusicMgr) + DepOut(VideoMgr)) = (1+2+1+1)+(0+2+0+0) = 7

Andre F. R. Cordeiro and Edson OliveiraJr / Journal of Computer Science 2018, 14 (3): 408.422 DOI: 10.3844/jcssp.2018.408.422



Fig. 3: MM2: Excerpt of the mobile media PLA

#### Metric for PLA Cohesion

HV *ariability*PLA counts the total relational cohesion (H) (Oizumi *et al.*, 2012) of components with variabilities. H is the average number of class internal relationships for a given component.

Equation 3 presents the Cohesion metric *HVariabilityPLA*:

$$HVariabilityPLA = \sum_{i=1}^{n} H$$
(3)

Where:

n = Number of components with variabilities

H = The relational cohesion of a component

Figure 3, FavouriteMgr has H 0 as class Favourite has H 0. MediaMgr has H 1 as class Media has H 1 and class MediaMgr has H 1. MusicMgr has H 1 as class MusicMgr has H 1 and class Music has H 1. VideoMgr has H 1 as class Video has H 1 and class VideoMgr has H 1. Thus HVariabilityPLA(MM2) = H(FavouriteMgr) + H(MediaMgr) + H(MusicMgr) + H(VideoMgr) = 1+1+1+1=4.

# **Experimental Validation of the Size Metric**

This section presents a controlled experiment for validating the metric for Size attribute, according to the template provided by Wohlin *et al.* (2000).

#### Definition

Considering the GQM template (Caldiera and Rombach, 1994), the experiment was performed to analyzing size metrics collected from UML models, for the purpose of validation, from the point of view of researchers, in the context of students in the role of SPL architects.

#### Planning

#### **Context Selection**

Academic constext was defined for this study.

#### Participants Selection

Selected participants were undergraduate, graduate students and lecturers from the State University of Maringa (UEM). For the selection of participants, non-probabilistic sampling was defined, for convenience. The 15 selected participants present the following profile: Graduate and undergraduate students, lecturers of Computer Science experts on modeling OO classes and components using UML and SPL, variability management and PLA. Experience is considered when one have modeled at least one OO system during one year in computer science course classes. SPL, variability management and PLA experience relies on designing PLA with variabilities in at least one system in any modeling language.

In this study, participants were not grouped. As the study considered metrics for the evaluation of PLAs, with no comparison between techniques and/or approaches, we decided to non-clustering them.

Selected participants have the following profile (Table 1):

- Undergraduate students from last semester of Computer Science and Masters and Ph.D. Candidates of Computer Science and Informatics
- All students with experience on modeling OO systems by means of at least class and component diagrams
- All students with knowledge on OO SPL, variability and PLA
- Researchers in Software Engineering from academia and industry with advanced knowledge on OO systems modeling, as well as SPL, variability and PLA

# Variables Selection

The following variables were defined for this experiment:

• Independent Variables: The PLA and the attribute to be evaluated in the experiment, in this case Size. PLA has a pre-fixed value (The Mobile Media PLA), whereas Size is a factor with two treatments, one referring to the metrics collected and one referring to the participants Size rating based on their experiences

• Dependent Variable: the correlation between treatments of the Size factor

#### Instrumentation

The instrumentation (Experimental package available at: http://www.din.uem.br/~edson/SMartyMetrics/jcs-paper) used in this experiment encompasses the following documents:

- Experimental Consent Form
- Characterization Questionnaire
- A Document with SPL basic concepts
- A Document describing the SMarty approach
- A document presenting the Size attribute and its metric
- A Document presenting the Mobile Media PLA
- An Evaluation Questionnaire on Size of the PLA evaluated

#### Hypotheses Formulation

The following hypotheses were defined for this study:

- Null Hypothesis (H0): There is no correlation between collected size metrics and the participants size rating based on their experience
- Alternative Hypothesis (H1): There is correlation between collected size metrics and the participants size rating based on their experience.

#### Experimental Design

We considered experimental and statistical test principles for designing this study, such as randomization at assigning PLAs to participants, nonclustering and non-balancing.

The principles of non-clustering and non-balancing were considered because of the characteristics of the study, which did not compare techniques and/or approaches.

The experimental arrangement considered was one factor and two treatments, in addition to a pre-fixed variable, PLA, whereas the Size factor had two treatments.

With relation to the statistical tests, the Spearman Correlation Test (Spearman, 1987) was applied correlates size metric collected to the size rate by participants experience.

#### Execution

#### Preparation and Tasks

Before starting the execution of the experiment, a training session was carried out lasting for about 50 min. Participants received information on SPL, variability management, PLA, size and reusability.

Participant	Education	Academia-industry	Experience	Experience in UML (classes)	Experience in SPL and Variability
1	Master	Academia	4 Years	Moderate	Supercial
2	Taking	Masters industry	2.5 Years	Moderate	Moderate
3	Master	Academia	4 Years	Moderate	Moderate
4	Taking masters	Academia	2 Years	Basic	Supercial
5	Taking masters	Academia	2 Years	Moderate	Basic
6	Taking masters	Academia	0.5 Year	Basic	Basic
7	Master	Academia	4 Years	Advanced	Advanced
8	Taking masters	Academia	5 Years	Moderate	Moderate
9	Taking Ph.D.	Academia	10 Years	Moderate	Supercial
10	Master	Academia	4 Years	Advanced	Advanced
11	Taking masters	Academia	5 Years	Moderate	Supercial
12	Master	Academia	7 Months	Basic	Supercial
13	Taking masters	Academia	3 Years	Basic	Basic
14	Taking masters	Academia	8 Months	Advanced	Advanced
15	Taking masters	Academia	7 years	Moderate	Moderate

#### Table 1: Size metrics experiment: Participants profile

Specifically on Size and Reusability, we presented definitions of both attributes. For Reusability, the following definition was presented: "Evaluation of the degree to which parts of the system can be used to build other systems" (Sommerville, 1985). With regard to Size, the following definition was presented: "number of operations of an element. This element can be of design (class models, interfaces, packages and components) or source code (representations in code of classes, interfaces, packages and components)" (Wust, 2017; SFPD, 2017).

After the training session, participants performed the following tasks:

- Sign the experimental consent term
- Complete the characterization questionnaire
- Read the document on SPL
- Read the document on SMarty
- Read the document on Size and Reusability attributes
- Read the document describing the Mobile Media PLA
- Respond to the Size Assessment Questionnaire

Reading tasks were carried out with the purpose of reinforcing the information presented in the training session. We believed such reinforcement improve the participants level of knowledge at the time of the PLA evaluation.

#### Executing the Experiment

The first activity was reading and signing the consent term. This term explains the purposes of the study, highlighting the privacy of personal information. Then, the characterization questionnaire was made available and each question was explained.

After completing the characterization questionnaire, documents describing SPL, PLA, variability management and Size and Reusability were made available to the participants.

The SPL used in the experiment was also explained in detail. Such an explanation considered the purpose of the SPL, as well as its main characteristics. It was explicitly mentioned that the PLA to be evaluated represented the architecture of such SPL.

Finally, the PLA to be evaluated and the evaluation questionnaires were made available and the experiment started. Participants were asked to mark the start and end time of the evaluation. During the experiment, the consultation of the documents was allowed.

All the material used by the participants was printed, from the supporting documents, to the PLA and the evaluation questionnaires.

#### Data Validation

The tasks performed by the participants were done and metrics were collected.

The evaluations of the participants were considered reliable for two reasons: the level of knowledge of the participants, as verified in the characterization questionnaire and the training performed, before the execution of the experiment.

#### Analysis and Interpretation

This section presents analysis and interpretation of results.

#### Preparation of Data Set

Data collected in the experiment was tabulated in spreadsheets. We organized the information from the characterization questionnaire, answered by the participants and the results of the evaluations (participant rates and metrics), for the correlation calculation. Basically two sets of data (participants and metrics) were used in the correlation calculation.

In addition to the tabulation, numerical conversion of the Likert scale labels for participants Size rates was necessary. The conversion consists of associating a numeric value for each label, according to Table 2.

With regard to the values calculated for the application of the metrics, no conversion was required. Therefore, the values presented are equal to the values collected.

#### Descriptive Statistics

As mentioned in Section 4.4.1, two sets of data were originated in the experiment. Descriptive statistics is shown for such data. Table 3 displays the two sets of data, as well as their mean, median and standard deviation.

Considering the participants size rating, we notice most of them classified the Reusability level of the PLAs as high or extremely high (values 1 and 2 respectively in Table 3, column \Participant Rate"). With regard to the values collected by the metrics, it is observed that most of them are in range 4 to 7, thus the majority of classes presented number of operations between 4 and 7.

#### Spearman Correlation Test

The Spearman Correlation Test (Spearman, 1987) is a non-parametric test based on the positions of the elements in two sets of data. According to the level of correlation observed a correlation indicator between the sets is obtained. Figure 4 presents the Spearman correlation scale. This non-parametric test was chosen as there was a conversion from Likert scale labels to numeric values (see Section 4.4.1).

By calculating the correlation we obtained the value of 0.3937237. According to Figure 4, this value represents a weak and positive correlation between

Table 3. Collected size 

metrics for Size and participants size rating for the evaluated PLAs. It means the judgment of participants with regard to size (reusability) is aligned to the metrics collected. Thus, considering this correlation value, we can reject H0 with 95% of assurance, thus providing initial evidence that the proposed size metrics can be used to assess the degree of reusability of PLAs.

Although this paper provides such initial evidence, two important aspects are observed.

First, new studies must be carried out as we cannot claim that the result obtained comes up only with the direct relation between Size and Reusability.

Second, Size might not be sufficiently adequate to evaluate PLA reusability as literature presents works, such as Her et al. (2007), that take into account the combination of different attributes to evaluate core assets reusability.

Table 2: Likert scale label conversion to numeric values

Reusability level	Numeric scale
Extremely low	-2
Low	-1
Neither low nor high	0
High	1
Extremely high	2

Partic. #	Size metric (nsopla)	Participant rate
1	7	2 (Extremelly high)
2	7	1 (High)
3	7	1 (High)
4	7	2 (Extremelly high)
5	5	1 (High)
6	5	2 (Extremelly high)
7	5	2 (Extremelly high)
8	5	1 (High)
9	4	1 (High)
10	4	-1 (Low)
11	4	1 (High)
12	4	1 (High)
13	4	2 (Extremelly high)
14	2	1 (High)
15	1	1 (High)
Mean	4.7333333	-
St. Dev.	1.791514	-
Median	5	1 (High)

-1	.0	Strong negative correlation	-0.5	Weak negative correlation	(	) Weak positive correlation	0.5	Strong positive correlation	1.0
Perfect correla	nega tion	tive			no cor	relation	I	Per cor	fect positive relation

Fig. 4: Spearman correlation scale (Spearman, 1987)

# **Experimental Validation of Coupling and Cohesion Metrics**

This section presents the experiment carried out to validate Coupling and Cohesion metrics.

# Definition

This experiment aim at analyzing metrics collected from UML models, with the purpose of validation, with respect to their capability of coupling and cohesion indicators, from the point of view of researchers, in the context of students and lecturers in the role of SPL architects.

# Planning

# Context Selection

This experiment context was academic.

# Participants Selection

Participants selected to this experiment have the same profile of the experiment of Section 4.2.2.

Table 4 presents information collected from the participants profile.

# Variables Selection

The following variables were defined:

- Independent Variables: The PLA, the attributes Coupling and Cohe-sion. The PLA is pre-fixed (Mobile Media PLA) and Coupling and Cohesion factors have two treatments each: One for the defined metrics and one for the participants rate. Such factors have no interaction
- Dependent Variables: the correlation between the metrics collected and the participants rate on coupling and cohesion

# Instrumentation

Instrumentation (Experimental package available at: http://www.din.uem.br/~edson/SMartyMetrics/jcs-paper) for this experiment includes most of the documents from Section 4.2.4, except the document describing the Size attribute, which was replaced with Coupling and Cohesion documents and the evaluation questionnaire.

#### Table 4: Coupling and cohesion experiment: Participants profile

# Hypotheses Formulation

We defined two sets of hypotheses for this experiment, one for coupling and one for cohesion, as follows:

- Coupling Null Hypothesis ((H0<sub>Cpl</sub>)): There is no correlation between collected coupling metrics and the participants coupling rating based on their experience
- Coupling Alternative Hypothesis (*H*1<sub>*Cpl*</sub>): There is no correlation between collected coupling metrics and the participants coupling rating based on their experience
- Cohesion Null Hypothesis (*H*0<sub>*Coh*</sub>): There is no correlation between collected cohesion metrics and the participants cohesion rating based on their experience
- Cohesion Alternative Hypothesis (*H*1<sub>*Coh*</sub>): There is a correlation between collected cohesion metrics and the participants cohesion rating based on their experience

# Experimental Design

For this experiment we chose a 2 factors with 2 treatments each as experimental design. However, the two factors do not interact to each other.

# Execution

# Preparation and Tasks

Before starting the experiment we carried out a training session. In this session, participants received information on SPL, variability management, PLA, Modularity, Coupling and Cohesion.

Definitions on Modularity, Coupling and Cohesion were presented to participants. For Modularity the following definition was presented: "it evaluates the degree to which the system is subdivided into cohesive logical parts, so that changes in one of these parts have minimal impact on the others".

With regard to Coupling, the following definition was presented: "Coupling is defined as the level of interdependence between different software modules, that is, the level at which different design units (classes, packages, components) are connected".

Participant	Education	Academia industry	Experience	UML (component (components)	Exp. SPL and variability
1	Taking masters	Academic	7 Years	Moderate	Basic
2	Taking masters	Academic	20 Years	Advanced	Advanced
3	Taking masters	Industry	3 Years	Moderate	Moderate
4	Taking masters	Academic	7 Years	Moderate	Moderate
5	Master	Academic	4 Years	Advanced	Advanced
6	Masters	Academic	4.5 Years	Advanced	Advanced
7	Taking masters	Academic	6 Years	Moderate	Moderate
8	Taking masters	Academic	1.2 Years	Basic	Basic

For Cohesion: "level at which internal elements of a design unit are logically related" (Myers, 1975; Chidamber and Kemerer, 1994; ISO/IEEE/IEC, 2010; Sommerville, 1985; SFPD, 2017; Wust, 2017). After the training session, participants performed the same tasks described in Section 4.3.1, except with relation to

documents for coupling and cohesion attributes.

#### Executing the Experiment

Execution of this experiment was the same as in Section 4.3.2, except this one considered coupling and cohesion rather than size.

#### Data Validation

The tasks performed by the participants were done and metrics were collected.

The evaluations of the participants were considered reliable for two reasons: The level of knowledge of the participants, as verified in the characterization questionnaire and the training performed, before the execution of the experiment.

#### Analysis and Interpretation

This section presents analysis and interpretation of results.

#### Preparation of Data Set

Data collected in the experiment was tabulated in spreadsheets. We organized the information from the characterization questionnaire, answered by the participants and the results of the evaluations (participant rates and metrics), for the correlation calculation. Basically two sets of data (participants and metrics) were used in the correlation calculation.

In addition to the tabulation, numerical conversion of the Likert scale labels for participants Coupling and Cohesion rates was necessary. The conversion consists of associating a numeric value for each label, according to Table 5.

#### **Descriptive Statistics**

Table 6 presents data and respective descriptive statistics for Coupling and Cohesion.

Most of participants rate coupling as high or low and the range of the metric DepCompV ariabilityPLA was 4 to 6 with median 0.5. For cohesion, most of participants rate it as high and the range of the metric HV ariability PLA was 5.66 to 6.66 with median 6.66.

#### Spearman Correlation Tests

We chose the Spearman Correlation Test due to a conversion from Likert scale labels to numeric values (see Section 5.4.1).

By calculating the correlation we obtained the value of 0.3443784 for coupling. According to Figure 4, this value represents a weak and positive correlation between metrics for Coupling and participants coupling rating for the evaluated PLAs. It means the judgment of participants with regard to coupling (modularity) is aligned to the metrics collected. Thus, considering this correlation value, we can reject H0Cpl with 95% of assurance, thus providing initial evidence that the proposed coupling metrics can be used to assess the degree of modularity of PLAs.

With regard to cohesion, we obtained the value -0,4206409 for correlation. This value represents a weak and negative correlation between metrics for Cohesion and participants cohesion rating for the evaluated PLAs. It means the judgment of participants with regard to cohesion (modularity) is aligned to the metrics collected. Negative correlation means that metrics collected and participants rate are inversely proportional.

 Table 5: Likert scale label conversion to numeric values

Modularity level	Numeric scale
Extremely low	-2
Low	-1
Neither low nor high	0
High	1
Extremely high	2

Table 6: Collected coupling an	d cohesion metrics and	participant rates
--------------------------------	------------------------	-------------------

	Coupling)	Part.	Cohesion	Part.
Partic. #	(DepCompVariabilityPLA	Coupling rate	(HVariabilityPLA)	Cohesion rate
1	5	1 (High)	6.66	1 (High)
2	5	2 (Extremelly High)	6.66	-2(Extremely Low)
3	4	1 (High)	5.66	1 (High)
4	5	-1 (Low)	6.66	1 (High)
5	4	-1 (Low)	5.66	1 (High)
6	6	1 (High)	7.66	0 (Neither Low nor High)
7	6	0 (Neither Low nor High)	7.66	1 (High)
8	4	-1 (Low)	5.66	1 (High)
Mean	4.875	-	6.535	-
St. Dev.	0.834523	-	0.834523	-
Median	5	0.5 (Neither Low nor High/High)	6.66	1 (High)

Thus, considering this correlation value, we can reject H0Coh with 95% of assurance, thus providing initial evidence that the proposed cohesion metrics can be used to assess the degree of modularity of PLAs.

Although this paper provides initial evidence on metrics for coupling and cohesion: New studies must be carried out as we cannot claim that the result obtained comes up only with the direct relation between Size and Reusability.

Coupling and Cohesion might not be sufficiently adequate to evaluate PLA modulatiry, such as Size.

# Threats to Validity and Lessons Learned

#### Conclusion Validity

The main threat in these experiments was the sample size. Several invited participants were not able to perform the experiments due to enrollments to masters and Ph.D. courses. However, as the purpose was to provide initial evidence on the existence of a correlation between metrics and participants rate, we understand that for generalizing such findings we need a large sample.

#### Internal Validity

The following issues might threat our experiments:

- Fatigue effect: To avoid fatigue, our training session for the size experiment lasted for 45 min and the experiment itself for about 30 min. For coupling and cohesion experiment, the session training lasted for 57 min and the experiment itself for about 44 min. In both experiments, participants rest for 30 min before start the experiment tasks
- Differences in participants knowledge: Participant profiles were different, thus to mitigate such threat we provided training sessions. Such sessions enable participants to get common knowledge on the subjects of interest in the experiments. Therefore, all participants were given the same training content
- Participant answers accuracy: We believe accuracy on participants answer is really hard to measure. However, we believe our instrumentation and training sessions helped to mitigate such threat. In addition, our instrumentation was previously evaluated for two lecturers/researchers of the SPL area

# Construct Validity

The questionnaires applied to participants in the experiments took into account a Likert scale with 5 labels. Such labels were converted to numerical values from -2 to +2 to perform correlations between metrics and participants rate. Such a conversion might threaten the construction of the experiments. Thus, to mitigate the inuence of such scales in the participants rating, we provided training sessions using Likert scales. Thus, we believe that a proper training might guide participants in choosing the best labels from such scales.

#### External Validity

One of the hardest tasks in software engineering experiments is recruiting participants. To mitigate such threat, we selected graduate students and lecturers from de SPL area. Most of selected participants had knowledge according to the required profile for our experiments. The ones who did not were closely observed during training sessions.

Another external threat to our experiments was the instruments used to evaluate the PLAs. Such PLAs came from the Mobile Media SPL, a pedagogical non-real SPL. We chose Mobile Media as its PLA is not too large in terms of classes and components. We believe that largest SPLs might threat the participants task realization due to excessive information represented. We understand this SPL might not threaten our experiments as several different works adopted it, such as (Colanzi and Vergilio, 2014; Oizumi *et al.*, 2012). However, we are aware that prospective experiments with larger and real SPLs must be considered.

#### Lessons Learned

Several lessons were learned with the conduction of our experiments.

Training is essential for reinforcing main concepts required for our experiments. Several participants from both experiments had sufficient knowledge on the subject of interest. However, we find training sessions an investment on gathering reliable evidence given the specific PLA and metrics validated. In our training sessions, SPL, PLA and variability management based on SMarty were revisited by multimedia educational resources, as well as exercises similar to the experiment tasks to memorize the training content. Brakes between training sessions and experiments tasks were also important due to fatigue effects. For experiments such as ours, we believe 30 min breaks are sufficient to mitigate threats during performing experiments task.

Pilot projects are straightforward to make experimental instrumentation more reliable and avoid misunderstanding instructions on experiments. In our case, one pilot project conducted before each experiment, was thus instrumentation was evaluated by an experienced lecturer on SPL, PLA and variability management. In our case, experiment on Size metrics benefit from pilot project as errors were found in the instrumentation. In addition, such evaluation enables state direct instructions for participants, which we widely encourage. A simple misunderstood instruction on an experiment task might threat data collected thus, consequently, data analysis and interpretation. Therefore, we encourage elaborate simple and objective instructions or questions to participants. If such instruction/question is complex, we suggest as the best strategy to split it into two or more parts, especially

for subjective/open questions, which depended on the participant interpretation. Using a non-real and small SPL (Mobile Media) benefit our experiments as: it has a small number of classes and interfaces, thus requiring participant less effort to count class/components elements; all SPL requirements are modeled in only one simple diagram (class and component); and the diagram is not too large for understanding it. We cannot say anything on these experiments using a real SPL, except that we expect more time to complete all tasks in prospective studies. The point is that real SPLs provide real data, thus, maybe contributing to findings similar to real projects.

# Conclusion

Considering the use of software metrics to represent attributes in a PLA, this paper presented the experimental validation of three metrics for PLAs. These metrics are associated with Size, Coupling and Cohesion attributes and unlike other size, coupling and cohesion metrics found in the literature, they are used in the context of SPL.

The metrics proposed and validated in this study were adapted from existing software metrics by taking into consideration PLA variabilities. Such adaptation occurred with the intention to provide SPL community a way of measuring PLA by means of external quality attributes, such as Reusability and Modularity based on Coupling, Cohesion and Size attributes.

The obtained results from our experiments provided initial evidence of the use of such metrics as indicators of Reusability (Size) and Modularity (Coupling and Cohesion) of a PLA. Reusability and Modularity were associated with Size and Coupling and Cohesion, respectively, taking into account works from (Chidamber and Kemerer, 1994; Myers, 1975).

As future work we intend to: (i) conduct more experiments with real SPLs and larger sample size to mitigate the threats discussed in this study; (ii) contribute to ISO/IEC 25010 standard (ISO/IEC,25010, 2011) in the perspective of Maintainability of PLAs supported by Modularity, Reusability, Analyzability, Modifiability and Testability by means of the validated metrics in this study and incorporating previously validated metrics for complexity (Marcolino *et al.*, 2013a) and extensibility (OliveiraJr and Gimenes, 2014); and (iii) conduct empirical evaluations of such ISO/IEC contribution with experts on software product quality and SPL.

# Acknowledgment

The authors would like to thank participants of the two experiments, who were available to evaluate PLA during experimental tasks and the lecturers attended the pilot projects of the experiments, thus improving instrumentation.

# **Funding Information**

Coordenação de Aperfeiçoamento de Pessoal de Nivel Superior (CAPES), a Brazilian funding agency, supported Andre F. R. Cordeiro with a 2-year masters scholarship.

# **Author's Contributions**

Andre F.R. Cordeiro: Coordinated the experiments performed at the Informatics Department of the State University of Maringá by planning, executing, analyzing and interpreting them. He also contributed in writing this paper.

**Edson OliveiraJr:** Supervised André during the experiments, contributed to planning, executing, analyzing and interpreting the experiments. He also contributed in writing this paper.

# Ethics

Experiments participants were previously invited, thus they attended such experiments as volunteers. Participants were informed on the confidentiality of personal data, as well as the evaluation will contribute to a non-profit research project coordinated by Dr. Edson OliveiraJr.

# References

- Bass, L., P. Clements and R. Kazman, 2012. Software Architecture in Practice. 3rd Edn., Addison-Wesley Professional.
- Bera, M.H.G., E. OliveiraJr and T.E. Colanzi, 2015. Evidence-based SMarty support for variability identification and representation in component models. Proceedings of the 17th International Conference on Enterprise Information Systems, Apri. 27-30, Barcelona, pp: 295-302. DOI: 10.5220/0005366402950302
- Bosch, J., R. Capilla and R. Hilliard, 2015. Trends in systems and software variability [Guest editors' introduction]. IEEE Software, 32: 44-51. DOI: 10.1109/MS.2015.74
- Caldiera, V. and H.D. Rombach, 1994. The goal question metric approach. Encycl. Software Eng., 2: 528-532
- Capilla, R., J. Bosch and K. Kang, 2013. Systems and software variability management: Concepts, Tools and Experiences, 1st Edn, Springer Science and Business Media, Heidelberg, ISBN-10: 3642365833, pp: 317.
- Chen, L., M.A. Babar and N. Ali, 2009. Variability management in software product lines: A systematic review. Proceedings of the 13th International Software Product Line Conference, Aug. 24-28, San Francisco, pp: 81-90.
- Chidamber, S.R. and C.F. Kemerer, 1994. A metrics suite for object oriented design. IEEE Trans. Software Eng., 20: 476-493. DOI: 10.1109/32.295895

- Colanzi, T.E. and S.R. Vergilio, 2014. A feature-driven crossover operator for product line architecture design optimization. Proceedings of the IEEE 38th Annual Computer Software and Applications Conference, Jul. 21-25, IEEE Xplore Press, Vasteras, pp: 43-52. DOI: 10.1109/COMPSAC.2014.11
- Colanzi, T.E., S.R. Vergilio, I. Gimenes and W.N. Oizumi, 2014. A search-based approach for software product line design. Proceedings of teh 18th International Software Product Line Conference, Sept. 15-19, Florence, Italy, pp: 237-241. DOI: 10.1145/2648511.2648537
- Fenton, N. and J. Bieman, 2014. Software Metrics: A Rigorous and Practical Approach. CRC Press, ISBN-10: 1439838232, pp: 617.
- Galster, M., D. Weyns, D. Tofan, B. Michalik and P. Avgeriou, 2014. Variability in software systems, a systematic literature review. IEEE Trans. Software Eng., 40: 282-306.
- Her, J.S., J.H. Kim, S.H. Oh, S.Y. Rhew and S.D. Kim, 2007. A framework for evaluating reusability of core asset in product line engineering. Information Software Technol., 49: 740-760.
- ISO/IEC,25010, 2011. Systems and software engineering - systems and software quality requirements and evaluation (SQuaRE) system and software quality models. https://www.iso.org/standard/35733.html.
- ISO/IEEE/IEC, 2010. Systems and software engineering. Vocabulary," ISO/IEC/IEEE 24765:2010(E)
- Kim, T., I.Y. Ko, S.W. Kang and D.H. Lee, 2008. Extending ATAM to assess product line architecture. Proceedings of the 8th IEEE International Conference on Computer and Information Technology, Jul. 8-11, IEEE Xplore Press, Sydney, pp: 790-797. DOI: 10.1109/CIT.2008.4594775
- Marcolino, A.S., E. OliveiraJr, I.M. Gimenes and E.F. Barbosa, 2017. Variability resolution and product configuration with smarty: An experimental study on UML class diagrams. J. Comput. Sci., 13: 211-226.
- Marcolino, A., E. OliveiraJr, I.M.S. Gimenes and T.U. Conte, 2013a. Towards validating complexity-based metrics for software product line architectures. Proceedings of the Brazilian Symposium on Software Components, Architectures and Reuse, Sept. 29-Oct. 4, IEEE Xplore Press, Brasilia, pp: 69-79. DOI: 10.1109/SBCARS.2013.18
- Marcolino, A., E. OliveiraJr, I.M.S. Gimenes and J.C. Maldonado, 2013b. Towards the effectiveness of a variability management approach at use case level. Proceedings of the 25th International Conference on Software Engineering and Knowledge Engineering, (EKE' 13), pp: 214-219.

- Montagud, S., S. Abrahao and E. Insfran, 2012. A systematic review of quality attributes and measures for software product lines. Software Qual. J., 20: 425-486.
- Myers, G., 1975. Reliable Software Through Composite Design. 1st Edn., Van Nostrand Reinhold, New York, ISBN-10: 0442256205, pp: 159
- Oizumi, W.N., A.C.C. Junior, G.G. Correia, T.E. Colanzi and S. Ferrari *et al.*, 2012. On the proactive design of product-line architectures with aspects: An exploratory study. Proceedings of the IEEE 36th Annual Computer Software and Applications Conference, Jul. 16-20, IEEE Xplore Press, Izmir, Turkey pp: 273-278. DOI: 10.1109/COMPSAC.2012.38
- OliveiraJr, E., I.M.S. Gimenes and J. C. Maldonado, 2008. A Metric Suite to Support Software Product Line Architecture Evaluation. Proceedings of the Latin American Computing Conference, (ACC' 08), pp: 489-498.
- OliveiraJr, E. and I.M.S. Gimenes, 2014. Empirical validation of product-line architecture extensibility metrics. Proceedings of the 16th International Conference on Enterprise Information Systems, Apri. 27-30, Lisbon, pp: 111-118. DOI: 10.5220/0004745201110118
- OliveiraJr, E., I.M.S. Gimenes, J.C. Maldonado, P.C. Masiero and L. Barroca, 2013. Systematic evaluation of software product line architectures. J. Universal Comput. Sci., 19: 25-52.
- OliveiraJr, E., J. Maldonado and I. Gimenes, 2010. Empirical validation of complexity and extensibility metrics for software product line architectures. Proceedings of the 4th Brazilian Symposium on Software Components, Architectures and Reuse, Sept. 27-29, IEEE Xplore Press, Bahia, pp: 31-40. DOI: 10.1109/SBCARS.2010.13
- OliveiraJr, E., I.M.S. Gimenes and J.C. Maldonado, 2010. Systematic management of variability in UML-based software product lines. J. Universal Comput. Sci., 16: 2374-2393.
- Olumofin, F.G. and V.B. Misic, 2005. Extending the ATAM architecture evaluation to product line architectures. Proceedings of the 5th Working IEEE/IFIP Conference on Software Architecture, Nov. 6-10, IEEE Xplore Press, Pittsburgh, pp: 45-56. DOI: 10.1109/WICSA.2005.33
- Osman, M.H., M.R.V. Chaudron and P. van der Putten, 2013. An analysis of machine learning algorithms for condensing reverse engineered class diagrams. Proceedings of the 29th IEEE International Conference on Software Maintenance, Sept. 22-28, IEEE Xplore Press, Eindhoven, pp: 140-149. DOI: 10.1109/ICSM.2013.25

Ribeiro, H.B.G., S.R.L. Meira, E.S. Almeida, D. Lucredio and A. Alvaro *et al.*, 2010. An assessment on technologies for implementing core assets in service-oriented product lines. Proceedings of the 4th th Brazilian Symposium on Software Components, Architectures and Reuse, Sept. 27-29, IEEE Xplore Press, Bahia, Brazil, pp: 90-99. DOI: 10.1109/SBCARS.2010.19

SDMetrics, 2017. Sdmetrics framework, list of metrics.

SFPD, 2017. Sdmetrics framework, properties of design.

Sommerville, I., 1985. Software Engineering. 2nd Edn.,

Addison-Wesley, Wokingham,

ISBN-10: 0201142295, pp: 334.

- Spearman, C., 1987. The proof and measurement of association between two things. Am. J. Psychol., 100: 441-471.
- Thurimella, A.K. and B. Bruegge, 2012. Issue-based Variability Management. Inf. Softw. Technol., 54: 933-950. DOI: 10.1016/j.infsof.2012.02.005

Wohlin, C., P. Runeson, M. Host, M.C. Ohlsson and B. Regnell *et al.*, 2000. Experimentation in Software Engineering: An Introduction. 1st Edn., Springer Science and Business Media, Boston, ISBN-10: 1461546257, pp: 204.

Wust, J., 2017. Sdmetrics: The software design metrics tool for uml.

# Appendix A - Size, Coupling and Cohesion Questionnaire

This appendix section provides Size, Coupling and Cohesion questionnaires.

Evaluation Instrument					
Start Time:	End Time:				
	Participant ID				

#### I. Context

This document presents questions, which must be answered. Resolution of such questions corresponds to the evaluation of Reusability of a Mobile Media variant PLA. Individual classes and interfaces of the PLA are evaluated with relation to reusability, then the overall PLA.

#### **II. PLA Evaluation**

For the given Mobile Media PLA, please evaluate the number of operations and the reusability of each class/interface.

Class / Interface	Is class/ interface present in the PLA?	What is the number of class/interface operations, discarding <u>get</u> and <u>set</u> methods and constructors?	Given the number of operations of a given class/interface, do you consider such class/interface easy to reuse?
Manager	[]Yes[]No		[]Yes[]No
AlbumManager	[]Yes[]No		[ ] Yes [ ] No
PhotoManager	[]Yes[]No		[ ] Yes [ ] No
Copyist	[]Yes[]No		[ ] Yes [ ] No
SmsTransfer	[]Yes[]No		[ ] Yes [ ] No
Sorter	[]Yes[]No		[ ] Yes [ ] No
Favourite	[]Yes[]No		[]Yes[]No
Photo	[]Yes[]No		[]Yes[]No
Media	[]Yes[]No		[] Yes [] No
Album	[]Yes[]No		[ ] Yes [ ] No

Considering the performed evaluation of all classes/interfaces, how do you rate the given PLA reusability?

Extremely Low	Low	Neither Low nor High	High	Extremely High



#### I. Context

This document presents questions, which must be answered. Resolution of such questions corresponds to the evaluation of Modularity of a Mobile Media variant PLA based on Coupling and Cohesion. Individual classes and interfaces of the PLA are evaluated with relation to modularity, then the overall PLA.

#### **II. PLA Evaluation**

For the given Mobile Media PLA, please evaluate the components internal and external relationships.

Component	Is component Present in the PLA	Does the component present <u>provided</u> <u>interfaces</u> used by other components?	Does the component present <u>required interfaces</u> that demand other components?	What kind of relationships the component presents the most:	Do you consider this component has more:
AlbumCtr1	[ ] Yes [ ] No	[ ] Yes [ ] No	[ ] Yes [ ] No	[] Internal [] External	[ ] Coupling [ ] Cohesion
AlbumMgr	[]Yes[]No	[ ] Yes [ ] No	[ ] Yes [ ] No	[ ] Internal [ ] External	[] Coupling [] Cohesion
MediaMgr	[]Yes[]No	[ ] Yes [ ] No	[]Yes[]No	[ ] Internal [ ] External	[] Coupling [] Cohesion
MusicMgr	[ ] Yes [ ] No	[ ] Yes [ ] No	[ ] Yes [ ] No	[ ] Internal [ ] External	[] Coupling [] Cohesion
PhotoMgr	[]Yes[]No	[ ] Yes [ ] No	[]Yes[]No	[] Internal [] External	[] Coupling [] Cohesion
VideoMgr	[]Yes[]No	[ ] Yes [ ] No	[ ] Yes [ ] No	[ ] Internal [ ] External	[] Coupling [] Cohesion
SenderMgr	[]Yes[]No	[ ] Yes [ ] No	[ ] Yes [ ] No	[ ] Internal [ ] External	[ ] Coupling [ ] Cohesion
LabelMgr	[ ] Yes [ ] No	[ ] Yes [ ] No	[ ] Yes [ ] No	[] Internal [] External	[] Coupling [] Cohesion
FavouriteMgr	[]Yes[]No	[]Yes[]No	[ ] Yes [ ] No	[ ] Internal [ ] External	[] Coupling [] Cohesion
Component	Is component Present in the PLA	Does the component present <u>provided</u> <u>interfaces</u> used by other components?	Does the component present <u>required interface</u> that demand other components?	What kind of relationships the component presents the most:	Do you consider this component he has more:
MediaCtrl	[ ] Yes [ ] No	[]Yes[]No	[ ] Yes [ ] No	[ ] Internal [ ] Externa	1 [] Coupling [] Cohesion
CopyMgr	[ ] Yes [ ] No	[ ] Yes [ ] No	[ ] Yes [ ] No	[] Internal [] Externa	1 [] Coupling [] Cohesion
SortMgr	[]Yes[]No	[]Yes[]No	[]Yes[]No	[] Internal [] Externa	1 [] Coupling [] Cohesion
EntryMgr	[ ] Yes [ ] No	[ ] Yes [ ] No	[]Yes[]No	[] Internal [] Externa	1 [] Coupling [] Cohesion
UserMgr	[ ] Yes [ ] No	[ ] Yes [ ] No	[]Yes[]No	[] Internal [] Externa	1 [] Coupling [] Cohesion

How do you rate the PLA coupling level?

Extremely Low	Low	Neither Low nor High	High	Extremely High
			~	
1				

How do you rate the PLA cohesion level?

Extremely Low	Low	Neither Low nor High	High	Extremely High

How do you rate the PLA modularity level?

Extremely Low	Low	Neither Low nor High	High	Extremely High