World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:12, No:3, 2018

Empirical Exploration of Correlations between Software Design Measures: A Replication Study

Jehad Al Dallal

Abstract**—**Software engineers apply different measures to quantify the quality of software design. These measures consider artifacts developed at low or high level software design phases. The results are used to point to design weaknesses and to indicate design points that have to be restructured. Understanding the relationship among the quality measures and among the design quality aspects considered by these measures is important to interpreting the impact of a measure for a quality aspect on other potentially related aspects. In addition, exploring the relationship between quality measures helps to explain the impact of different quality measures on external quality aspects, such as reliability and maintainability. In this paper, we report a replication study that empirically explores the correlation between six well known and commonly applied design quality measures. These measures consider several quality aspects, including complexity, cohesion, coupling, and inheritance. The results indicate that inheritance measures are weakly correlated to other measures, whereas complexity, coupling, and cohesion measures are mostly strongly correlated.

Keywords**—**Quality attribute, quality measure, software design quality, spearman correlation.

I. Introduction

SOFTWARE engineers do not only aim to develop functionalities required for software systems, they also aim to ensure that developed systems are of high quality. Software users typically avoid using systems of low quality because such systems potentially have low maintainability, reliability, and usability.

Researchers have proposed several measures to quantify or estimate different internal and external quality attributes (e.g., [1]-[5]). Internal quality attributes, such as cohesion, coupling, and complexity, can be quantified using software artifacts. External quality attributes, such as reliability and maintainability, consider software environment artifacts, such as numbers of detected faults or numbers of changed lines of code during software maintenance phases. Although external quality attributes are of interest for software users, they cannot be measured during software development phases because they consider artifacts that are unknown during these stages. To overcome this problem, researchers have worked on building statistical models that use software artifacts considered by internal quality attributes to predict external quality attributes.

Cohesion and coupling are key internal quality attributes. Cohesion refers to the relatedness between module components, and coupling refers to the relatedness between software modules [6]. Typically, software engineers aim to develop systems with modules of high cohesion and low coupling. However, these two quality attributes can potentially impact each other greatly, therefore software engineers have to keep a balance between them, as improving one attribute potentially weakens the other.

In general, empirically exploring the relationship between internal quality measures provides interpretations for the impact of improving one quality aspect on other aspects. In addition, such empirical investigation helps researchers to explain the effects of different internal quality attributes alongside external quality attributes. Understanding the relationship between the internal quality measures helps software engineers and practitioners to determine the quality factors that have to be considered and focused on during quality assessment processes.

In this paper, we empirically explore the correlation between six design quality attributes known as CK measures [2]. These measures have been widely applied in practice and have been considered through research. These measures consider several internal quality attributes, including complexity, cohesion, coupling, and inheritance. Although several studies have empirically investigated the relationship between these measures (e.g., [7]-[10]), more replication studies are required to approve or disapprove the reported findings. The study reported in this paper involved three software systems selected from different domains.

The reported results show that cohesion has a negative impact on complexity, coupling, and inheritance, and that complexity, inheritance, and coupling have positive impacts on each other. We found that the correlation between the measures of inheritance and other quality aspects were weak, whereas the correlation between measures of coupling, cohesion, and complexity were moderate to strong.

This paper is organized as follows. Section II provides an overview of the CK measures and the correlation results reported in previous studies. Section III describes the considered systems and the data collection process. Section IV reports the correlation study analysis and discusses its results. Finally, Section V concludes the paper and discusses future work.

II. Related Work

Researchers have proposed many measures to quantify or estimate software quality attributes and have studied them both theoretically and empirically (e.g., [1]-[5], [10]-[12]). The most studied and applied measures are those proposed by

Jehad Al Dallal is with the Department of Information Sciences, Kuwait University, P.O. Box 5969, Safat 13060, Kuwait (e-mail: j.aldallal@ku.edu.kw).

Chidamber and Kemerer [2]. The CK metric suite includes the six measures that their definitions are included in Table I. These measures quantify four quality aspects, including complexity, cohesion, coupling, and inheritance. WMC indicates complexity, DIT and NOC indicate inheritance, RFC and CBO indicate coupling, and LCOM indicates absence of cohesion. Although these measures are considered in many studies, few of them reported the correlation coefficients between the measures.

TABLE I
DEFINITIONS OF THE CK MEASURES [13]

Measure	Definition
Weighted Methods per Class (WMC)	Number of methods defined in a class.
Depth of Inheritance Tree (DIT)	Number of ancestors of a class.
Response for a class (RFC)	Number of methods that are either defined in class A or directly invoked by the methods in class A.
Number of Children (NOC)	Number of classes that directly inherit class A.
Lack of Cohesion Methods (LCOM)	Number of pairs of methods that do not access common attributes.
Coupling Between Object classes (CBO)	Number of classes that access methods or attributes of class A + Number of classes that are accessed by the methods or attributes of class A. Classes that are inherited by class A are excluded.

Basili et al. [6] considered student systems, including 180 C++ classes, and used Pearson's correlation technique to obtain the correlation coefficients between the CK measures. The results indicated that the correlations between CK measures were, in general, very weak. Gyimothy et al. [7] considered a total of 3,192 classes in two C++ systems, applied Pearson's correlation technique, and found that the correlations between the measures were weak for one system and mostly weak to moderate for the other system. Shatnawi and Li [8] considered several releases of Eclipse and three CK measures including CBO, RFC, and WMC. They applied Spearman's correlation technique to empirically explore the correlation between the considered measures. The results indicate a strong correlation between the three considered CK measures. Finally, Singh et al. [9] considered 145 C++ classes, and their results indicated a strong correlation between RFC and both WMC and DIT. All other pairs exhibited weak correlations.

III. SELECTED SYSTEMS

We selected three open source Java software systems from three different domains: Eclipse v.1.0.1 [14], Art of Illusion v.2.5.1 [15], and DrJava v.beta-20090505-r4932 [16]. Eclipse was released on September 8, 2009 and included 1,438 concrete classes (as opposed to abstract classes or interfaces). Art of Illusion was released on October 22, 2007 and consisted of 471 concrete classes. DrJava was released on May 5, 2009 and consisted of 1,036 concrete classes. These systems were randomly selected from http://sourceforge.net. We used the CKJM tool [17] to obtain the required quality values. Tables II-IV show the descriptive statistics for each of the selected measures, including the minimum, mean, maximum, and standard deviations for each of the three selected systems.

TABLE II
DESCRIPTIVE STATISTICS OF THE CK MEASURES FOR ECLIPSE'S CLASSES

Measure	Minimum	Maximum	Mean	Std. Dev.
WMC	0	341	15.60	24.89
DIT	0	8	2.04	1.22
NOC	0	33	0.45	1.60
CBO	0	451	15.17	28.43
RFC	0	661	41.11	64.29
LCOM	0	57520	381.73	2357.08

TABLE III
DESCRIPTIVE STATISTICS OF THE CK MEASURES FOR ART OF ILLUSION'S CLASSES

Measure	Minimum	Maximum	Mean	Std. Dev.
WMC	1	110	13.65	13.35
DIT	0	9	2.15	1.72
NOC	0	55	0.45	3.01
CBO	0	224	23.48	27.39
RFC	1	612	51.14	56.00
LCOM	0	3459	79.95	278.99

TABLE IV
DESCRIPTIVE STATISTICS OF THE CK MEASURES FOR DRJAVA'S CLASSES

Measure	Minimum	Maximum	Mean	Std. Dev.
WMC	0	796	14.30	37.90
DIT	1	9	2.54	1.69
NOC	0	20	0.49	1.82
CBO	0	447	13.29	23.16
RFC	0	1526	33.39	75.72
LCOM	0	226338	671.78	8741.22

IV. CORRELATION STUDY ANALYSIS AND RESULTS

We applied the nonparametric Spearman correlation statistical technique to obtain the correlation coefficients between the six considered measures. We selected this technique because the obtained quality values were found to not follow the normal distribution. The correlation coefficient values within the ranges [0, 0.2[, [0.2, 0.4[, [0.4, 0.6[, [0.6, 0.8[, and [0.8, 1] were respectively classified as very weak, weak, moderate, strong, and very strong. We reported only the statistically significant (p-value < 0.0001) results in Tables V-VII for the three considered systems.

The results reported in Tables V-VII led to the following observations:

1. WMC had statistically significant correlation to each of the other measures. Except for one case, WMC had a positive correlation to all of the other measures. This indicated that complexity has a positive correlation to inheritance and coupling and a negative correlation to cohesion. WMC has at least a strong correlation to both RFC and LCOM, a mostly strong correlation to CBO, and at most a weak correlation to both DIT and NOC. These observations indicate that complexity attribute has at most a weak correlation to inheritance and at least a strong correlation to both the coupling and cohesion attributes.

TABLE V
CORRELATION COEFFICIENTS OF THE CK MEASURES FOR ECLIPSE'S CLASSES

Measure	WMC	DIT	NOC	CBO	RFC	LCOM
WMC	1.00	0.06	0.31	0.69	0.91	0.88
DIT	0.06	1.00	-	0.11	0.17	0.06
NOC	0.31	-	1.00	0.36	0.27	0.33
CBO	0.69	0.11	0.36	1.00	0.75	0.64
RFC	0.91	0.17	0.27	0.75	1.00	0.78
LCOM	0.88	0.06	0.33	0.64	0.78	1.00

TABLE VI
CORRELATION COEFFICIENTS OF THE CK MEASURES FOR ART OF ILLUSION'S CLASSES

Measure	WMC	DIT	NOC	CBO	RFC	LCOM
WMC	1.00	0.235	0.150	0.618	0.762	0.679
DIT	0.235	1.00	-	0.252	0.426	0.113
NOC	0.150	-	1.00	0.193	-	0.213
CBO	0.618	0.252	0.193	1.00	0.684	0.473
RFC	0.762	0.426	-	0.684	1.00	0.518
LCOM	0.679	0.113	0.213	0.473	0.518	1.00

TABLE VII
CORRELATION COEFFICIENTS OF THE CK MEASURES FOR DRJAVA'S CLASSES

Measure	WMC	DIT	NOC	CBO	RFC	LCOM
WMC	1.00	-0.09	0.15	0.48	0.87	0.76
DIT	-0.09	1.00	-	-	-	-
NOC	0.15	-	1.00	0.21	0.12	0.12
CBO	0.48	-	0.21	1.00	0.52	0.48
RFC	0.87	-	0.12	0.52	1.00	0.63
LCOM	0.76	-	0.12	0.48	0.63	1.00

2. In most cases, DIT had statistically significant correlation to each of the other measures, except NOC. Considering the statistically significant results and, except for one case, DIT had a positive correlation to each of the other measures. This indicates that inheritance has a positive correlation to complexity and coupling and a negative correlation to cohesion. Except for one case, DIT had at most a weak correlation to WMC, CBO, RFC, and LCOM, indicating that inheritance has at most a weak correlation to complexity, coupling, and cohesion.
3. In most cases, NOC had a statistically significant correlation to each other measure except DIT. NOC had a positive correlation to WMC, CBO, RFC, and LCOM. Similar to the observation reported for DIT, NOC's results indicated that inheritance had a positive correlation to complexity and coupling and a negative correlation to cohesion. NOC had at most a weak correlation to WMC, CBO, RFC, and LCOM, indicating that inheritance attribute had at most a weak correlation to complexity, coupling, and cohesion. This observation was similar to that reported for DIT.
4. Except for one case, CBO had a statistically significant correlation to each other measure. CBO had a positive correlation to each of the other measures. This indicated that coupling had a positive correlation to inheritance and complexity and a negative correlation to cohesion. For most cases, CBO had a strong correlation to WMC and RFC, a moderate correlation to LCOM, and at most a weak correlation to DIT and NOC. These observations indicated that coupling had at most a weak correlation to inheritance, a moderate correlation to cohesion, and a strong correlation to complexity.
5. Except for two cases, RFC had a statistically significant correlation to each of the other measures. RFC has a positive correlation to each of the other measures. This indicates that coupling attribute has a positive correlation to inheritance and complexity and a negative correlation to cohesion. Based on majority (i.e., vote counting), RFC has a very strong correlation to WMC, a strong correlation to each of CBO and LCOM, and at most a weak correlation to each of DIT and NOC. These observations indicate that coupling attribute has at most a weak correlation to inheritance, and at least a strong correlation to each of complexity and cohesion.
6. Except for one case, LCOM had a statistically significant correlation to each of the other measures. It has a positive correlation to each of the other measures. This indicates that cohesion has a negative correlation to inheritance, complexity, and coupling. Based on majority (i.e., vote counting), LCOM had a strong correlation to both WMC and RFC, a moderate correlation to CBO, and at most a weak correlation to both DIT and NOC. These observations indicate that cohesion has at most a weak correlation to inheritance, a moderate to strong correlation to coupling, and at least a strong correlation to complexity.

Table VIII summarizes the observations indicated by the obtained results, and the signs between parentheses indicate the direction of impact.

TABLE VIII
SUMMARY OF THE OBSERVATIONS

Quality attribute	Complexity	Inheritance	Coupling	Cohesion
Complexity	-	Weak (+)	Strong (+)	Strong (-)
Inheritance	Weak (+)	-	Weak (+)	Weak (-)
Coupling	Strong (+)	Weak (+)	-	Moderate/ Strong (-)
Cohesion	Strong (-)	Weak (-)	Moderate/ Strong (-)	-

V. CONCLUSIONS AND FUTURE WORK

This paper explores the correlation between CK software design measures. Analyzing the correlation between these measures helps indicate the correlation between four design quality attributes, including complexity, inheritance, cohesion and coupling. The study involved classes of three open source Java systems. Key observations indicate that cohesion has a negative correlation to each of the other attributes and inheritance, complexity, and coupling have positive correlations to each other. In addition, the results indicate that complexity, coupling, and cohesion have moderate to strong correlation between each other and weak correlations to inheritance.

The obtained results and observations were different from those reported in [6], [7], and [9], which were based on

Pearson's technique or other unspecified correlation techniques. We believe that the results reported in this paper are more reliable because they are based on more classes and systems and use a statistical technique that is more suitable for the values of quality measures, which typically do not follow normal distribution. On the other hand, the results reported in this paper align with those reported in [8].

In the future, we plan to perform more replication studies that involve more systems, measures, and quality attributes.

References

[1] R. Jabangwe, J. Börstler, D. Šmite, C. Wohlin, Empirical evidence on the link between object-oriented measures and external quality attributes: a systematic literature review, Empirical Software Engineering, 2015, 20(3), pp 640-693.
[2] Chidamber, S. R. and Kemerer, C. F., A Metrics suite for object Oriented Design, IEEE Transactions on Software Engineering, Vol. 20, No. 6, 1994, pp. 476-493.
[3] J. Al Dallal and L. Briand, A precise method-method interaction-based cohesion measure for object-oriented classes, ACM Transactions on Software Engineering and Methodology (TOSEM), 2012, Vol. 21, No. 2, pp. 8:1-8:34.
[4] J. Al Dallal and L. Briand, An object-oriented high-level design-based class cohesion metric, Information and Software Technology, 2010, 52(12), pp. 1346-1361.
[5] J. Al Dallal, Fault prediction and the discriminative powers of connectivity-based object-oriented class cohesion metrics, Information and Software Technology, 2012a, 54(4), pp. 396-416.
[6] V. Basili, L. Briand, W. Melo, A validation of object-oriented design metrics as quality indicators, IEEE Transactions on Software Engineering, 1996, 22(10), pp. 751-761.
[7] T. Gyimothy, R. Ferenc, and I. Siket, Empirical validation of object-oriented metrics on open source software for fault prediction, IEEE Transactions on Software Engineering, 2005, 3(10), pp. 897-910.
[8] R. Shatnawi and W. Li, The effectiveness of software metrics in identifying error-prone classes in post-release software evolution process, The Journal of Systems and Software, 2008, 81, pp. 1868-1882.
[9] Y. Singh, A. Kaur and R. Malhotra, Empirical validation of object-oriented metrics for predicting fault proneness models, Software Quality Journal, 2010, 18, pp. 3-35.
[10] J. Al Dallal, Incorporating transitive relations in low-level design-based class cohesion measurement, Software: Practice and Experience, 2013, Vol. 43. No. 6, pp. 685-704.
[11] J. Al Dallal, Transitive-based object-oriented lack-of-cohesion measure, Procedia Computer Science, Volume 3, 2011, pp. 1581-1587.
[12] J. Al Dallal, Accounting for data encapsulation in the measurement of object-oriented class cohesion, Journal of Software: Evolution and Process (Wiley), Vol. 27, No. 5, 2015, pp. 373-400.
[13] J. Al Dallal and S. Morasca, Investigating the impact of fault data completeness over time on predicting class fault-proneness, submitted for publication in Information and Software Technology, 2017.
[14] Eclipse, http://www.eclipse.org/, accessed in March 2017.
[15] Illusion, http://sourceforge.net/projects/aoi/, March 2017.
[16] DrJava, http://sourceforge.net/projects/drjava/, accessed in March 2017.
[17] CKJM extended - An extended version of Tool for Calculating Chidamber and Kemerer Java Metrics (and many other metrics), http://gromit.iiar.pwr.wroc.pl/p_inf/ckjm/, accessed in January 2017.