An Exploration of Object Oriented Metrics in Student Code

Stijn Mijland

Oof, you just lost a lot of progress. That's a deep frustration, a real punch in the gut. - Bennett Foddy

Contents

Chapter 1

Introduction

For the course Object Oriented Programming at the Rijksuniversiteit Groningen, students are tasked to work in pairs on 3 different programming projects over the duration of the course to become familiar with object orientation. When they are done, and when they have handed in their programs, those programs are evaluated by teaching assistants that give these projects a grade, and provide these students with feedback. In this manner, it tries to teach them object orientation.

The field of object oriented metrics attempts to classify object orientation in certain ways, and one of those ways is through object oriented metrics. These metrics, which attempt to quantify certain design qualities in object oriented source code, are an approximation of the qualities that are considered important in object oriented source code.

By approximating the quality of a project both via human means, through a grade, and through computational means, via object oriented metrics, every project ends up with two quality measurements, where one requires human input and the other does not. Consequently, the question arises how these quality indicators relate to one another. Is it perhaps possible to use these object oriented metrics to replicate the work of the teaching assistant?

There are advantages to automated grading. One would expect an automated grading system to produce results more quickly, to do so more consistently, and to do so more cheaply (although, as always, that depends on the approach). Moreover, besides the practical advantages, there are theoretical advantages, since humans are very capable of grading, but computers at time of writing are not.

At this time there is no known relation between object oriented metrics and a grade a teaching assistant gives. This thesis aims to figure out if such a relation between object oriented metrics and a grade given by a teaching assistant can be established. While this is by no means an automated grading system, it will establish if object oriented metrics may serve as a basis or component in building such a system.

Besides looking at the relation between object oriented metrics and grades, this thesis looks at the effects that the groups of students have on both grades and metrics, and does the same for each of the 3 projects students worked on, as well as the teaching assistants that graded them. These effects may bias the data, making them relevant for those that would design an automated grading system for object oriented programs.

1.1 Gathered Data

This thesis concerns itself with code produced during a course in Object Oriented Programming, which means that it is limited by the data that is gathered during that course. Moreover, this research is not concurrent with said course, reducing the possibility of gathering more, potentially interesting data impractical. Therefore, the scope of this thesis is itself limited by the data gathered during said Object Oriented Programming course. For this reason it will be easier to understand the research questions for this thesis by first explaining the available data.

1.1.1 Assignments

While following the course on Object Oriented Programming, students will encounter 3 graded assignments. They are provided with an introduction assignment that familiarizes them with the way the course is ran, as well as introducing some of the basics of using the programming language they will be using during the course, which in this case is the Java programming language. However, that assignment is only there to try and eliminate operational difficulties during the assignments, which is why no data is gathered about this introduction. Data is gathered about the 3 graded assignments only. These assignments are called RPG, Card Game and Graph Editor. An overview:

RPG Students are tasked with creating a Role Playing Game with a hierarchy of characters and rooms. Characters inherit interaction options, such that students become familiar with how interacting with a general character will have different behaviour depending on their concrete type. The same is true for Rooms, which can do things upon entry, exit, and inspection. In this manner, the concepts of state, behaviour, and polymorphism through inheritance are made apparent.

Card Game Card Game was made to make students familiar with writing graphical user interfaces with the Java Swing framework. In earlier versions of the course, this was done during Graph Editor, but Graph Editor did not show students how to work with Swing. The Card Game assignment provides a basic card game program, and asks the student to implement their own game with it. This allows them to familiarize themselves with the Swing framework without the need to dig through tutorials to find what they need. It also gives them a chance to expand on other people's code, which will be useful for later programming projects.

Graph Editor Graph Editor is an assignment that asks students to create a program that deals with visual graphs, and the ability to edit them. It involves saving and writing files, implementing Model-View-Controller to show and manipulate it, and asks students to do this from the ground up. Just like card game it involves a graphical user interface, but unlike card game students will be organizing all of the code themselves.

1.1.2 Groups

For this course, students work on these assignments in groups of about 2 students. In order to accommodate an uneven number of students, as well as the possibility that students drop out of the course, some groups of 1 person or 3 people are allowed. This does mean that a single group is not necessarily a specific group of people, but in general this is the case. Each group is identified by number, and there are about 80 groups in this course in total. Groups were assigned a number in order of application, i.e. the first group to apply received group number 1.

It should be noted that some students were members of different groups while this course took place. However, that information has not been registered. For this reason this thesis will assume that a group number means an exact combination of specific students. It will also assume that these groups are operating independently. This is perhaps not entirely consistent with the true nature of university work, but the alternative is constantly observing students when they work on an assignment.

It is not the intention of this work to implement mass surveillance on students. Every group will have at most 3 observations. A low amount of observations is much more impactful in reducing the chances of finding measurable and valid effects. Independence of observations is indeed important, but that will already be the case for many groups, as not all groups interact with each other, and it is not sensible, especially not for a study this size, to invade privacy to such a degree to guarantee independence of groups when other solutions are at least equally effective in improving the quality of results.

1.1.3 Teaching Assistants

Because 80 groups all work on the same assignment, and all have the same deadline, and because the course intends to provide feedback within 24 hours, it is not feasible to grade everything within a working day. Therefore, to spread the grading workload, there are teaching assistants assigned to the course that work on a portion of the grading. For this course there were 3 pairs of teaching assistants and one group of 3 teaching assistants. To protect their privacy, these teaching assistants are also referred to by number.

It should be noted that the group of 3 teaching assistants worked in pairs on some projects, and worked alone on others. This means that there are multiple configurations of this group that must be accounted for. This will be discussed in the methodology section.

To divide the grading workload, teaching assistants were divided over groups of students equally (in a round robin fashion). When groups stopped participating in the course that meant their teaching assistants had less to do than the other groups of teaching assistants. For this reason the groups were slightly rebalanced after every assignment to maintain a fair work load. Because of this, student groups will indeed have some correlation with teaching assistant groups, but because it was intentional, there is nothing to discover in analysing that.

Teaching assistants are not strictly speaking independent as they share a metric and are able to discuss grading details with each other. However, this was mostly done within the groups the teaching assistants worked in, and not outside of those groups, making these groups mostly independent, just like with the groups of students.

1.1.4 Grades

The teaching assistants gave a grade to each project. This data is included in the data for this thesis. This is a grade between 0 and 10, where 10 is the best value. In some cases (such as a late submission) students received a late penalty, which reduced the grade. When that happens, the grade is no longer only dependent on the code itself. To attempt to mitigate that problem, the grade without those subtractions is also included.

Although the grade is arrived at via a rubric, the elements of that rubric are not general to all projects. Most of the elements are about specific functionality. This means that most of those elements can not be used when analysing all projects at the same time. Of the rubric elements, two of them, OO Standard and Coding Standard are kept for analysis. It must be mentioned however that the Graph Editor assignment does not have these data points.

Leaving out data points that are not common between all projects makes it easier to compare the projects, but when it comes to designing a grading system including this extra data would have made sense. For this project the comparability has been chosen over the specificity, because this project looks at the general effects of the metrics on grades. Follow-up research attempting to create a more specific automated grading system should consider analysing the full rubric.

1.1.5 Conclusion

This section described which data was gathered from the Object Oriented Programming course that it is based on. In summary, this is the Groups students worked in, the Projects they worked on, the Teaching Assistants that graded the assignments, and the grade they received for the assignment. This is not all available data, because the names of both students and teaching assistants, and the contents of the grading rubric have been excluded.

Because this thesis decided to focus on properties that can be generalized across projects, and because this thesis decides to respect the privacy of the students and teaching assistants it bases its research on, that data will indeed not be included. That means that the available data is the Groups of students, the Assignment they made, the Teaching Assistants that graded them and the Grade they received, along with the Corrected Grade, OO Standard and Coding Standard.

1.2 Research Questions

The available data for this thesis consists of Groups of students, the Projects they worked on, the Teaching Assistants that graded them, and the Grades they gave for those projects. Along with that data, the intention is to calculate object oriented Metrics. These five properties, Groups, Projects, Teaching Assistants, Grades and Metrics comprise the data that can be analysed.

Groups, Projects and Teaching Assistants are related variables, but only insofar that they are all part of the observation. Certainly, Groups were assigned a specific group of Teaching Assistants, but this was done to distribute work. There is no relation to be discovered, it was made this way. The same is true in the relations with Projects: both Groups and Teaching Assistants are involved in all of the Projects unless groups drop out. There is no relation to discover.

That only leaves the relations between grades and metrics to look at, both in direct relation, and in relation to the three categories they can be related to (Group, Project Teaching Assistant). The matter of exploring these relations is described in the following research questions:

Is there an effect of groups on project grades?

Is there an effect of groups on project metrics?

Is there an effect of teaching assistants on project grades?

Is there an effect of teaching assistants on project metrics?

Is there an effect of assignments on project grades?

Is there an effect of assignments on project metrics?

Is it possible to explain grades by means of metrics?

In this manner, the possible relations between these variables should be sufficiently explored in this thesis to give insight in how these variables behave, explaining to those that would look further into for example automated grading if this way of trying to explain grades works, and whether or not they can base their system on this, or if they must investigate another approach.

1.3 Related work

There is little that can be referenced for this section. Most of the relevant articles are discussed at length in the theoretical framework. During literature review some articles were found that use object oriented metrics, and even Alves rankings, but they are applied to production code, and specifically the detection of fault-proneness - a common topic in object oriented metric research.

One recent example of such research is (Boucher & Badri, 2017), who use a neural network to attempt to predict if a class is error prone. Because of the small amount of data in the data set used for this project, and because where the set used by (Boucher & Badri, 2017) is marked for errors, whereas this data set is not, the approach used in that paper is not applicable, but it is relevant as another application of the same techniques used in this project.

When referring to techniques, it is important to mention (Alves, Ypma, & Visser, 2010), which contains a comparison of various methods of aggregating data to project level. Many of those methods required data that was not available, which is why the Alves ranking described in (Alves et al., 2010) are used instead, but with different data sets those different methods may be useful.

As far as metrics are concerned, (Designite, 2019) is a useful research for calculating metrics, (Chidamber & Kemerer, 1994) is a widely used metrics suite, and many sources also consider (Abreu & Melo, 1996) for object oriented metrics, sources those intending to do similar research should find useful.

1.4 Theoretical Framework - Object Oriented Metrics

This thesis plans to use object oriented metrics in various analyses, which means that it should be expected that the results of those analyses reflect what is known about those object oriented metrics. To prepare for this this section looks at the definitions of object oriented metrics, and at research that has managed to discover relations between object oriented metrics and certain properties of source code.

1.4.1 Lines of Code

When it comes to metrics, the most commonly seen metric is the number of lines of code, but it is not an object oriented metric, because it applies to all programming languages that write their source code in text files, and does not measure any object oriented concept. Lines of code (or as it is given here: source lines of code, referring to the number of lines that change the behaviour of their program) is defined as follows:

SLOC - Source Lines of Code The lines of code an entity encompasses. What does or doesn't constitute a line is often up to the language standard, but generally speaking it refers to a series of characters followed by a newline or return carriage symbol, and excludes comment lines.

The reason Lines of Code is still mentioned here as a metric, is because it is straightforward to calculate, and because it is a trivial indicator of problems. When a piece of code becomes complex, it needs an equivalent amount of code to solve that complexity. As a result, it will gather a large number of lines. Moreover, if code has more lines, there is a higher chance that the author made a mistake in it. This makes lines of code a good indicator for both maintenance effort and risk of mistakes.

1.4.2 Chidamber & Kemerer

The first specifically object oriented metrics suite is A Suite for Object Oriented Metrics by Chidamber & Kemerer. This metric suite introduces and evaluates multiple object oriented metrics. These are the following metrics: (Chidamber & Kemerer, 1994)

WMC - Weighted Methods per Class The weighted methods per class metric is the sum of the complexity of all methods: $\text{WMC} = \sum_{i=1}^{n} c_i$ where c_i is the complexity of the i'th method. Those familiar with the subject might read this as cyclomatic complexity, but there is no standard definition to allow for alternative measures of complexity.

Having many methods, especially complicated methods make a class difficult to understand. Not all complexity can be avoided, but having a lot of complexity in one place is a maintenance risk. Moreover, many methods means that there are many places where the state of the object can be changed, which results in an increased risk of coding mistakes.

DIT - Depth of Inheritance Tree When a class A has extending classes $B_1, B_2, \ldots, B_N$, where N is the number of children, the DIT for that class A is

$$\text{DIT}(A) = \begin{cases} 0 & \text{if } N = 0 \\ 1 + \text{MAX}_{i=1}^{N}(\text{DIT}(B_i)) & \text{if } N > 0 \end{cases}$$

Having a deep inheritance tree means that changes in that class echo through many other classes in the system, potentially breaking a lot of code. This means that coding mistakes in classes with a high depth of inheritance tree propagate to a large part of the program, making these classes a priority target for testing effort.

Another aspect of deep inheritance tree is that while they can be extended, they reduce the maintenance effort. When the abstraction of the inheritance tree is no longer sufficient, and needs to be refactored, the entire tree must be replaced. The advantage of a deep inheritance tree is that many classes reuse the same code. The disadvantage is that when requirements change to make the existing design inappropriate, there is a large maintenance effort.

In effect, a deep inheritance tree for code with fixed requirements means a reduced testing effort, whereas a deep inheritance tree with changing requirements indicates a large amount of technical debt. This means that this metric is ambiguous to interpret without requirements information.

NOC - Number of Children The number of classes that directly inherit from a class C.

This metric is mostly similar to depth of inheritance tree, but in a more localized manner. Rather than indicating general reuse, this metric indicates the direct reuse of the class. While Depth of Inheritance Tree only scales with many levels of inheritance, this more closely indicates how much of the data is reused compared to Depth of Inheritance Tree. The authors treat them as related.

CBO - Coupling Between Object classes Often expanded as coupling between objects, instead describes the coupling of classes. If a method of one class C uses methods or variables of another class D, that is one coupling from C to D.

Having highly coupled classes is another indicator of changes in one class affecting a larger part of the system, which is a maintenance risk and makes class reuse more difficult (as coupled classes must also be imported).

RFC - Response For a Class The number of methods of a class, plus the methods called by each method. For a class A, let M_A be the methods of A. Let M_* be the methods of all classes in the system. Let the call set for a method be defined as $C(m) = \{n \in M_* | m \text{ calls } n\}$. Then the RFC is defined as $RFC(A) = M_A \cup \{C(m) | m \in M_A\}$

When a class uses a large amount of methods, it depends on a lot of code elsewhere in the system, meaning that changes elsewhere are likely to require change in this class, or breaking its functionality.

LCOM - Lack of Cohesion Of Methods The number of methods that don't use the same variable. Let M_A be the methods for class A. Let $V(m)$ be the set of variables used by method m, then

$$\text{LCOM(A)} = |\{(i,j) | i \in M_A \wedge j \in M_A \wedge i \neq j \wedge V(i) \cup V(j) = \emptyset\}|$$

Having incoherent methods indicates that a class uses functionality that could perhaps be split, producing classes that are individually easier to understand.

In experiments it has been found that Depth of Inheritance Tree, Response For a Class, and Coupling Between Objects are somewhat positively correlated with faults in classes. Number of Children on the other hand tends to be inversely related with faults. (Basili, Briand, & Melo, 1996)

1.4.3 MOOD - Metrics for Object Oriented Design

The MOOD metrics are metrics that attempt to quantify good or bad design. In contrast to the Chidamber & Kemerer metrics suite, these metrics are calculated at project level. The metrics defined in MOOD are the following metrics:

MHF - Method Hiding Factor Effectively the number of classes that can see the method of another class, but as a factor. Equals the number of hidden methods divided by the defined methods. This may be offset for the number of classes a method is visible to, to account for different access modifiers, such as protected and package.

Formally: given a method m, define $V(m)$ as the percentage of classes that can see m, excluding its declaring class. This means that if A extends B, then it can see B's protected methods, so that's more classes that can see it, despite the method being protected. The $V(m)$ of a private method is by corollary 0, and that of a public method is 1.

Given a class C, define M_d as the number of methods it declares, including inherited methods. Define M_v (visible) as $\sum_{i=1}^{N} V(m_i)$, where N is the number of methods, and m_i is the ith method. Define M_h (hidden) as $M_d - M_v$.

MHF is defined as $\frac{\sum_{i=1}^{T} CM_h(C_i)}{\sum_{i=1}^{T} CM_d(C_i)}$ where TC is the total number of classes.

By quantifying the ration of hidden methods, this metric quantifies to what degree the workings of classes can be influenced by other classes. This makes this metric an indicator of modularity, as well as an indicator for potential errors, assuming that all points of interaction may cause errors.

AHF - Attribute Hiding Factor Diverting from the term of field, calling it an attribute instead, this is identical to the MHF, except over the fields, rather than the methods, of the system.

Even more so than methods, public attributes are causes of errors, as most classes rely on class methods to control their state, and having external code modify those attributes means that those values must be checked every time they are used. Hence, a high attribute hiding factor indicates encapsulation, modularity, as well as a lower risk of modification errors.

MIF - Method Inheritance Factor The total methods compared to the declared methods. $MHC = \frac{\sum_{i=1}^{T} CM_i(C_i)}{\sum_{i=1}^{T} CM_a(C_i)}$ where M_i is $M_a - M_d$, M_a is the available methods (inherited and declared), and M_d is the methods declared in the class itself.

Inherited methods are methods defined in a parent class. This means that with respect to the properties of that parent class, they are expected to be consistent, and if they are, having many inherited methods reduces the risk that errors are created in each class.

Unfortunately, the reverse is also true. If many methods are inherited whilst containing errors, that means those errors also apply to the classes that inherit them. Moreover, the inheriting classes are able to interact with all the inherited method, multiplying the number of interactions and thus errors.

In the philosophy of object orientation inheritance is considered a positive because it allows reuse of code, thereby reducing the amount of locations in which errors might occur. However, many languages prefer composition over inheritance precisely because it reduces interactions, while still enabling code reuse.

At the very least, a high method inheritance factor indicates use of inheritance, reuse of code and consciousness of the availability of inheritance. Whether or not that is positive is up for debate, both in theoretical and practical terms.

AIF - Attribute Inheritance Factor Like MIF, but with methods.

Attribute inheritance factor is also related to use of inheritance, in much the same manner as method hiding factor is. However, it is often the case that an attribute and the methods to interact with it are not designed to be changed by other methods. This makes allow attribute inheritance factor arguably more important for reducing errors, no matter how efficient or useful the ability to change attributes from parent classes may be.

POF - Polymorphism Factor The degree of overriding relative to the amount of possible children. $PF = \frac{\sum_{i=1}^{T} CM_o(C_i)}{\sum_{i=1}^{T} CM_n(C_i) \cdot DC(C_i))}$ where M_o is the number of overridden methods for a class, M_n is the new methods it defines, and DC is the number of descendant classes.

Another mainstay of object orientation is the dynamic dispatch mechanism that allows code to interact with a generic class, while using the methods of a specific class. It is considered to be a way of writing generic code, as well as a mark of good object oriented design.

The problem with polymorphism, however, is that dynamic dispatch is inefficient, that its contract, in the shape of a method signature, is often not quite clear enough, leading to problems with interpretation, and that it is often applied when other approaches would have been just as good or even better, all for the sake of object orientation.

To believers of object orientation, Polymorphism Factor is an indicator of good design. To others, it can also be a cause of errors or slowdown, and a method of alternation that could very well be replaced with something easier to understand or maintain.

COF - Coupling Factor $CF = \frac{\sum_{i=1}^{T} C \sum_{j=1}^{c} (C_i, C_j)}{TC^2 - TC - 2 \cdot \sum_{i=1}^{T} CDC(C_i))}$ where $c(a, b)$ is 1 if a accesses fields and methods of b, and 0 otherwise.

Coupling factor is an indicator of the interaction between pieces of code. A high coupling factor implies that changes in one place will cause changes in another place, which means a higher maintenance risk. Low coupling can therefore be seen as a general positive for all code. However, some code must simply interact, so there is limited benefit in reducing coupling for the sake of reducing coupling.

Besides the intended benefits, these metrics, too, have been researched with respect to other benefits they might provide or data they might indicate. Specifically, their authors research their correlation with the defect density (the number of coding errors in a file), the failure density (how often the code in question failed), and the normalized rework effort (how many hours needed to be spent to make the code work as intended. They gathered these values from 8 different projects.

In that research, it seems that all these metrics are of influence on these parameters. Unfortunately, the significance of these results is not presented, and the way the metrics are used makes it seem like these results could well have come from overfitting. These metrics are relevant, but for this thesis their experimental results will be ignored. (Abreu & Melo, 1996)

1.4.4 Other Metrics Suites

Research also found the existence of a metric suite by Lorenz and Kidd (Bansal & Agrawal, 2014). This suite was also published in 1994, and is sometimes found in other literature. It does not seem to have been validated with respect to its relation to fault-proneness or other properties that would be interesting in general maintenance.

1.5 Methodology

This thesis revolves around seven research questions:

Is there an effect of groups on project grades?

Is there an effect of groups on project metrics?

Is there an effect of teaching assistants on project grades?

Is there an effect of teaching assistants on project metrics?

Is there an effect of assignments on project grades?

Is there an effect of assignments on project metrics?

Is it possible to explain grades by means of metrics?

This chapter describes how it plans to answer these questions. In general, two categories can be created using these questions: Questions that deal with categorical data, and questions that deal with continuous (or, to be precise, approximately continuous) data. The first 6 research questions fall in the first category, the final question falls in the last category.

1.6 Metrics and Statistics

Before any research questions can be answered, one problem needs to be solved. In this thesis, the intention is to compare multiple object oriented projects with each other using object oriented metrics. Certainly, these metrics can be calculated, and if they apply to entire projects, they can be compared directly. However, when it comes to the Chidamber & Kemerer metric suite, those metrics apply to classes. Some metrics even apply to methods. In both those cases, they do not apply directly at project level.

One way to deal with this problem, is to do nothing, and report every entity and their metric. The problem here is that that approach is very awkward. In the projects used for this thesis, students were allowed to design their own programs, which means that even classes with the same name can not necessarily be compared directly, as they may do different things.

Because of this, a table containing unaggregated data will have a column for each method and each class, containing exactly one value: the metric for that entity. Using such data in an analysis would involve aggregation regardless. Therefore, the preferable approach is to perform this aggregation before doing the analysis.

1.6.1 Aggregation of Metric Data

One way to approach the aggregation of metric data is to see each project of a collection of entity metrics, and to treat those metrics as numbers. Some straightforward ways to turn a series of numbers into one summarizing number is by looking at the minimum, the maximum, the mean, the median, the variance and the standard deviation.

Those aggregation operations are straightforward to implement, but they are also somewhat agnostic towards the nature of the data they are aggregating, which may result in less useful results. Therefore it seemed worthwhile to investigate other methods of aggregating metrics.

(Alves et al., 2010) lists a number of them before introducing their own approach, which is what is used here.

1.6.2 Alves Rankings

Alves rankings are a way of combining multiple software projects into one large data set, and then comparing them in terms of the amount of risk relative to the other projects. To do this, it only needs the data from these projects, making it suitable for use in situations without a quality reference. This is exactly the situation in this thesis.

To apply Alves rankings, one collects entities in software with an associated metric. For example, a record of a class and its number of lines of code. These records are collected and combined, such that there are records for every entity in every project. As an example, consider the following student project, which has 21 classes and 2 metrics: lines of code and lack of cohesion of methods:

Package Name	**Type Name**	**LOC**	**LCOM**	**Weight**	**Risk Factor**
nl.rug.oop.rpg	GameRegistry	37	0.500	0.029	Low Risk
nl.rug.oop.rpg	Properties	129	0.333	0.102	Low Risk
nl.rug.oop.rpg	Main	151	0.250	0.119	Low Risk
nl.rug.oop.rpg	MenuHandler	196	0.000	0.155	Low Risk
nl.rug.oop.rpg	Inspectable	27	0.000	0.021	Low Risk
nl.rug.oop.rpg	DescriptionNotSetException	14	0.000	0.011	Low Risk
nl.rug.oop.rpg	Player	99	0.313	0.078	Low Risk
nl.rug.oop.rpg	RandomGenerator	59	0.000	0.047	Low Risk
nl.rug.oop.rpg.items	DefencePotion	25	0.400	0.020	Low Risk
nl.rug.oop.rpg.items	Item	24	0.667	0.019	Very High Risk
nl.rug.oop.rpg.items	HeartContainer	18	1.000	0.014	Very High Risk
nl.rug.oop.rpg.items	HealthPotion	25	0.400	0.020	Low Risk
nl.rug.oop.rpg.npcs	Healer	47	1.000	0.037	Very High Risk
nl.rug.oop.rpg.npcs	NPC	11	1.000	0.009	Very High Risk
nl.rug.oop.rpg.npcs	Enemy	103	0.636	0.081	High Risk
nl.rug.oop.rpg.doors	CreepyDoor	34	0.667	0.027	Very High Risk
nl.rug.oop.rpg.doors	SpikyDoor	30	0.667	0.024	Very High Risk
nl.rug.oop.rpg.doors	Door	91	0.600	0.072	Medium Risk
nl.rug.oop.rpg.dataholders	PlayerData	13	0.000	0.010	Low Risk
nl.rug.oop.rpg.rooms	LootRoom	36	1.000	0.028	Very High Risk
nl.rug.oop.rpg.rooms	Room	98	0.500	0.077	Low Risk
Total	21	1267	-	1	

After that, each individual record is given a percentage that represents the percentage of its metric relative to the whole project. In our example, the project has a total of 1267 lines of code, and dividing each lines of code data point by 1267 produces the weight for each class, as can be seen in the table.

After assigning weights, a histogram of individual metric values in the project is produced. This histogram will have counts for each metric value measured (0, 0.25, 0.313, 0.333, 0.4, 0.5, 0.6, 0.636, 0.667, and 1 in the example). However, instead of counting each occurrence of a value once, the way a histogram is generally constructed, the amount counted is equal to the weight assigned in the first step. This means that smaller entities influence the calculation in proportion to their size. The result is the following histogram:

Values	Weighted Count	Cumulative	Risk
0.000	0.24	0.24	Low Risk
0.250	0.12	0.36	Low Risk
0.313	0.08	0.44	Low Risk
0.333	0.10	0.54	Low Risk
0.400	0.04	0.58	Low Risk
0.500	0.11	0.69	Low Risk
0.600	0.07	0.76	Medium Risk
0.636	0.08	0.84	High Risk
0.667	0.07	0.91	Very High Risk
1.000	0.09	1.00	Very High Risk

This is an example, and because it is an example, it has omitted an important step. This is the step where multiple histograms of metrics are combined into a single histogram to allow the comparison of projects. The combining of the histograms takes place after each metric value has received its weighted count. This combination of histograms takes place before calculating the cumulative weights.

Using the histograms with weighted metric values, a large histogram is made containing all measured metric values. The bins for these metrics values are calculated by summing all bins with the same metric value in the original histogram and dividing by the number of projects combined. Because all bins combined represent 100% of their source code, combining them in this manner keeps the histogram total at 100%. After this, the calculation proceeds as shown above, except that risk factors must be attributed to all projects involved.

The use of lines of code to generate weighted count means that a weight depends on the total lines of code for one project, not the total lines of code for all projects. This decision is a trade-off, which means it functions for projects of different size rather than just projects of similar size, at the cost of having smaller projects have a large influence on the overall result.

Because these histograms display how each metric is proportionally represented in its source, they are treated as having equivalent weight, which is why they are combined into a single histogram. This is done by joining all bins of all histograms, where any duplicate metric values have their values added, then dividing by the number of projects that are being evaluated.

Once the final histogram has been calculated, each metric value will have received a weighted count. A known property of object oriented metrics is that lower values are considered better. Therefore, starting at the smallest recorded metric value, a cumulative sum of percentages is calculated, as in the example. Using this cumulative value, every metric value is assigned a risk indicator according to the cumulative sum of percentages p. These values are taken directly from (Alves et al., 2010).

- Anything with $p < 70\%$ is marked as Low Risk
- $70\% <= p < 80\%$ is marked as Moderate Risk
- $80\% <= p < 90\%$ is marked as High Risk
- $p >= 90\%\%$ is marked as Very High Risk

Finally, in the original metrics overview, the metric value for each entity is looked up in the histogram, and the associated risk value is recorded, as can be seen in the example. In this manner Alves rankings can assign a risk factor to each entity, which is relative to the risk of all code combined, allowing it to highlight problematic code, irrespective of the average level of the code it is found in.

The open question then is how this information can be reported on the project level, because the reporting only reaches risk levels on the entity level, method or type. To aggregate these values on the project level, the simplest way is to count the number of entities of Low Risk, Moderate Risk, High Risk and Very High Risk level respectively. In this thesis, counting each level is the approach that is used. (Alves et al., 2010)

1.6.3 Percentile Method

Due to a misinterpretation of Alves rankings, data has also been calculated through the use of a simplified version of Alves rankings. This simplified method does not weight the classes using lines of code. Instead, it creates a histogram in the usual manner, by counting classes with a certain metric value. It then assigns risk factors in the same manner as Alves rankings, by calculating cumulative percentiles of classes represented. Risk assessments are also performed in the exact same manner.

The main advantage of this approach is that it can be calculated without knowing the lines of code for an entity. This makes this method independent for each metric. The main disadvantage is that by missing the weighting classes with 10 lines and classes with 1000 lines account, that differ by a factor 100 in size account for the same weight in the method, which means that many small classes can easily obscure large classes, making it less suitable for detecting issues.

This method is kept around because it may be interesting to see how this different approach performs, on its own, and in comparison to Alves rankings, If it never shows up, it probably was not very useful, and then the results can simply be ignored (at the expense of page space). If it does show up, it is sensible to research the method on its own as apparently it has merit.

1.7 Categories for Questions with Categorical Data

As mentioned, the first 6 research questions deal with categorical data. These data are: Groups, Teaching Assistants, and Projects. Before anything else, there is first the matter of how those are subdivided into categories. The data itself has been discussed before in the section on gathered data, where it was explained that each group received a number, and that each assignment has its own name. The categories for teaching assistants remained mostly untouched.

The reason for that is that the groups of teaching assistants were, during the course, referred to by their names. For privacy reasons, that can not be repeated here. Since some teaching assistants worked alone, and some together, the most sensible approach is to give each unit (a group of 1 or 2 teaching assistants that always worked together) a letter and proceed from there. This lettering is done as follows:

TA teams A, B and C always worked together. TA's D, E and F switched whom they worked with. To use this in analyses, this category must be consolidated to one category. What follows are three proposed methods to consolidate the TA category:

- Version A: the *first* letter in the alphabet among both TA categories becomes the category, i.e. if D and E worked together, the category becomes D.
- Version B: the *last* letter in the alphabet among both TA categories becomes the category. If D and E work together, under B the category becomes E.

- Version C: the combination of two TA's gains its own proper letter. The combination: $D+E$ becomes G, $D+F$ becomes H, $E+F$ becomes I.

Normally, categorization C would be the best, however, the data set is not very large for the single TA's, and thus, there is a risk that given the larger degree of freedom, the available data will not be enough to tell if there is an effect. Hence, the use of versions A and B, which, despite being less accurate, may be able to discover effects better. These three categories are the third, fourth, and fifth category in the data respectively.

Altogether this results in 5 categorizations: The Group, which will be a number between 1 to 81, inclusive; The Project, which will be the name of the specific assignment, and finally, the three categorizations of Teaching Assistants described above. These are the categories that will be used for the first 6 research questions.

1.8 Answering Questions with Categorical Data

All questions with categorical data concern themselves with both metrics and grades. What both metrics and grades have in common, is that they are mostly continuous data. Certainly, data like OO Standard only seems to have steps of 0.2 or 0.25 points on a range between 0 and 1, and grades and metrics are both susceptible to similar issues that means they are not exactly continuous, but for this thesis they are treated as if they were continuous. The reason this is important, is because it enables the use of ANOVA.

1.8.1 ANOVA

ANOVA, or ANalysis Of VAriance is a technique to look at data for samples that fall in different categories, and see if there is a difference in that data for these categories. For example, in medical trials, when one group has a placebo, and another group has a real treatment, ANOVA can be used to see whether or not there is an effect from the treatment. The real effect size can then be determined later with alternative methods.

In order to model these different categories, the intended result is to see if the means for individual categories differ. Therefore, each category is best modelled as a category mean. In order to model all values in multiple categories with their category means, they all need to receive their own error term for the model to be accurate. This leads to the following model:

Let $i \in [1, k]$ be the category [1] a data point is a member off. Let $j \in [1, n_i]$ be the jth sample in category i. Then each data point is written as x_{ij}. Let μ_i be the mean of category i. Let ϵ_{ij} be the error term for the value x_{ij}. Then the model for ANOVA is:

$$x_{ij} = \mu_i + \epsilon_{ij}$$

The goal is to see how likely it is that such means were found by chance, and if there is little chance it was accidentally discovered, consider there to be an effect, which is accomplished via the F-statistic, which will not be repeated here, as this thesis does not implement its own statistic. (Verzani, 2014a) What it does do, however, is consider the potential shortcomings

[1] kategory

of this test. As mentioned, the goal of ANOVA is to represent categories by a sample mean. Three important consequences of that approach are the following:

- The samples must be independent. If two or more means depend on the same value, their values will not have the difference they would have had if they were independent, which reduces the measured difference, and thereby the effectiveness of the test.
- The residuals must be normally distributed. The error terms ϵ_{ij} when plotted in ascending order should match a normal distribution. If they are not, that means that it is not valid to model this data using single means. Since finding a mean for each group is the goal of this model, not having normally distributed residuals reduces the effectiveness of the test.
- The variance must be homogeneous across categories. If one mean has a higher variance than another mean, the distribution of likely true values of that higher variance mean is wider than for the mean with less variance. Therefore, having differing variance between categories makes it harder to compare the means effectively, reducing the effectiveness of the test.

For this thesis, it is assumed that samples are made independently. Based on the way the data was gathered, this is not strictly guaranteed, hence this assumption. For the normality of residuals, the Shapiro-Wilk Normality Test is used, which is a statistical test where the p-value indicates the likelihood that the data used is normal. For the homogeneity of variance assumption, Levene's test is used, where the p-value indicates the likelihood that the variances are equal.

With this one assumption and two test, it would seem that it is trivial to determine whether or not an ANOVA is valid. However, this is not the case. This is because tests for normality of residuals and for homogeneity of variance are sensitive to large amounts of data. When the number of samples grows, they become very sensitive to small deviations, effectively punishing a use of larger sample sizes.

Because of this, the advice tends to be to observe plots of the relevant quantity (such as QQ-plots for normality of residuals, and boxplots for homogeneity of variance) instead of using these tests. Although these tests are effective at indicating assumption violations, they are not very useful to indicate the gravity of the violation.

1.8.2 ANOVA in This Thesis

In this thesis, the plan is to evaluate these questions regarding categorical data using One-Way ANOVAs, where the categories are the 5 categories specified before, and the analysed data are the various columns of grade data and metric data. The intention is to have a large amount of metric data (the exact amount does not belong in this methodology section) which means that there will be an even larger amount of analyses performed.

What's more: each of those ANOVAs is performed twice. This is because some student groups left the course before they were done and did not finish all assignments. Because of this, there are some groups that completed all assignments and some that did not. Because of this using all data outright introduces a certain bias towards students that left, which means that some effects may be amplified or dulled down. To make sure that those changes can be recognized, all ANOVAs are performed with all data, as well with data of students that completed all assignments.

In order to work around the amount of analyses, and the limited usefulness of having a significant result, the ANOVA results are included, but not in direct reports. Instead, they appear in large tables of ANOVA results, that are formatted as follows:

Variable	**Group P-value**	**Project P-value**	**TA's (A) P-value**	**TA's (B) P-value**	**TA's (C) P-value**
Data Group					
Data	p-value	p-value	p-value	p-value	p-value

These tables tabulate the results of performing ANOVA using data column Data, which is part of Data Group, with the category it crosses with, reporting the p-value of that ANOVA. For example, in the Data Group 'Resultant Variable', there is the data column 'Corrected Grade'. In these tables, each p-value can be marked with one of the following coloured labels:

Significance level Assumption	[1,0.1)	$[0.1, 0.05)$	$[0.05, 0.01)$	$[0.01, 0.001)$	$[0.001, 0)$
All Assumptions Met					
Homogeneous, not Normal					
No Assumptions Met					

These labels are intended to distinguish between different levels of significance for ANOVAs, as well as differentiating between which assumptions an ANOVA meets according to the Shapiro-Wilk normality test (indicated as 'Normal' above) and Levene's test for homogeneity of variance ('Homogeneous') To this end a colour is added to each cell to indicate the assumptions met, the opacity of the colour relates to the level of significance.

The astute reader will notice that there is no colour option for Normal, not Homogeneous. To arrive at this approach, some tests with the data were performed, and the results of those tests showed that such a situation does not occur in the data used in this thesis. What's more, this approach ignores the advice to look at plots rather than relying on tests.

While it is true that the Shapiro test and Levene's test have shortcomings when calculated on large amounts of data compared to looking at plots, looking at plots is manual intervention, whereas these tests can be performed automatically. During development of the above table it became clear that there were about 1000 ANOVAs. 1000 ANOVAs is not an unmanageable amount, but it is very likely that categorizing 1000 plots manually will introduce errors.

The ideal solution would be to use different tests that are less sensitive to larger data sizes, but creating such a test makes the conclusions drawn using such a test weak statistically, because the properties of that test have not been established in proof. Therefore, to maintain statistical validity and transparency while also trying to gather as many results as possible, all results are kept, but so are the tested assumption violations.

Using ANOVAs provides insight in the existence of effects from categories on data. However, it does not explain what the size of that effect is, how significant the effect is, and therefore what influence each individual factor concretely has on the result. For example, if there is a difference between how Teaching Assistants grade, then it is important to know where that difference can be found. Knowing that a difference exist is in itself not very useful.

To arrive at an in-depth answer to the research questions, it is therefore important to continue to investigate the difference between the factors in each category. However, the found differences are themselves a valid result for answering the research questions. For this reason, these results are included in this thesis. However, the next section will only be using significant ANOVAs ($p < 0.05$).

1.8.3 Filtering ANOVA results

Because of how p-values work (they indicate the chance that the result was caused by chance), when doing many analyses, you expect at least p percent to turn out significant (i.e. with $p = 0.05$ nearly 5 percent of tests should end up significant). Because of this, if a category does not produce at least p percent significant results, there is less information in that category than in random data, and it can be discarded.

Now, strictly speaking, the goal of every research question this thesis treats is to see if there is an effect of certain categories on metrics and grades. However, a more complete answer is not only able to assert that such an effect exist, but also the size of that effect. Before that can be done, however, it is necessary to ascertain that in general, it is valid to assume an effect exists.

This, in part is accomplished by either sanity checks for the data used in ANOVA, such as having a variance not equal to 0. The other part is by applying the property of likelihood inherent in the p-value as explained above.

To do this, the number of ANOVAs, and the amount of the ANOVAs that is significant is counted, and calculated as a percentage. This is then repeated for each of the colours that indicate which assumptions are met, just like in the full tables with ANOVA results. Through this table, a meta-analysis of ANOVAs is created that makes it possible to evaluate if in general these analyses can provide information.

In this manner, it is possible to filter the results for those analyses that are likely to contain information, which provides a useful basis for investigating further towards the difference between category effects.

To calculate the differences between factors for a significant ANOVA, Tukey Honest Significant Differences (TukeyHSD for short) is used. Tukey HSD calculates the difference between individual factors of a category, and reports a confidence interval for each difference (in the form of a mean, a lower bound and an upper bound), along with a p-value, which will allow the extraction of differences between individual factors in categories.

1.8.4 Tukey Honest Significant Differences

In the model of ANOVA, it was explained that ANOVA uses a sample mean for each category for its model. Consequently, the most straightforward way to calculate the size of the effect is to make pairwise subtractions of those means to see what the difference between means is. However, more information is available in the data.

TukeyHSD does look at the pairwise differences, but rather than producing just a difference value, gives an estimated difference, lower and upper bounds for the difference in a confidence interval that may be specified, and a p-value on the likelihood that the real difference is 0. This means that using TukeyHSD provides the difference in effect between groups, with an arbitrary desired level of confidence, as well as testing if that difference is non-zero.

1.8.5 Usage of TukeyHSD

Before performing TukeyHSD, all ANOVAs that are not significant are ignored. For each of those ANOVAs, TukeyHSD is calculated. If there are significant differences, each of the four values provided by TukeyHSD (lower bound, upper bound, estimate and p-value) is reported. The data is again colour coded and saturated in the way ANOVA is, using the Shapiro and Levene's test for colour and the Tukey p-value for saturation.

In testing for this tabulation it showed that there are very few significant results for groups (details will be given in the relevant section later), too few to be considered a valid source of information. A table with 3000 parallel entries that is readable on A4 paper would be a non-trivial data formatting problem, but for this thesis that problem has proven irrelevant.

The end result is a number of tables that indicate which differences between groups are significant and what the calculated difference is. This should provide a more thorough analysis of what the difference between categories means for various data. Again, testing showed that this will still be a large amount of data to handle, which is why here too, a meta analysis is in order.

Here too, it is possible to tabulate the amount of significant differences versus the total amount of differences measured, allowing the use of the properties of the p-value to eliminate certain results. This meta-analysis takes place in much the same manner as with ANOVA, reporting on all 3 levels of met assumptions.

After this meta-analysis, there is little that can be done to automate analysis still, which is why from that point the approach will be to look at the produced tables and see which results stand out, and report those. This may produce a haphazard conclusion, because it only looks at statistically confirmed differences, not at general differences, which means that a general conclusion may be difficult to make. However, such a conclusion can then be drawn using ANOVA, where those results can serve as an example.

It should be noted here that to accommodate the interpretation of ANOVA results for TukeyHSD, the risk factors for Alves rankings and the Percentile method are divided by their totals. This is because when calculating these methods for different projects, they report the amount of entities in each risk category. Because of this, if two projects have 2 Very High Risk classes, there is no measured difference, yet when one project has 100 classes, and the other only 50, that 50-class project is clearly worse than the other.

That same calculation is trivial to interpret if each risk category is divided by the total. In that case, the comparison would not be between 2 and 2, but between 0.02 and 0.04, making the 50-class project clearly worse than the 100 class project, and also making it so that significant differences are interpretable.

1.9 Grades and Metrics - Linear Models

There is one research question remaining, which is about the relationship between grades and metrics. To analyse that relation, it is interesting to look at the nature of the data. Reiterating from the collected data:

OO Standard	A quality value between 0 and 1, inclusive describing the quality of object-orientation in a project
Coding Standard	A quality value between 0 and 1, inclusive describing the degree to which coding practices (such as comments, indentation) were applied in a project
Grade	The final grade given to a project on a scale from 0 to 10, with 10 being the best grade.
Corrected Grade	The final grade without any detractions (such as late penalties) included on the same scale as the grade

These values are grades, which means they must reflect how well a student did an assignment, and while there is a limit to the precision that can be exercised in determining a grade, the intention is for the grades to be continuous in their domain. Therefore statistics that deal with continuous values are appropriate.

1.9.1 Linear Regression

With linear regression, the goal is to find a linear equation that using the data available gives the lowest error in its dependent variable. There are many ways of representing this, but in this representation, y' is the predicted value for the dependent variable y, X is a matrix of observed data, and β represents the model parameters.

$$y'_i = \beta_0 + X_{i1}\beta_1 + X_{i2}\beta_2 + \cdots + X_{iN}\beta_N$$

Linear regression then tries to find values of β such that $\epsilon = \sum_{i=1}^{N}(y'_i - y_i)^2$ is minimized. That way, the result is the linear model that best explains the variable as a direct linear result of the data. (Verzani, 2014b) With this model, it is possible to see how accurately the data was approximated, and therefore, how accurately metrics are able to predict grades.

One challenge of a linear model is the selection of parameters. To do this, the R programming language offers a function called 'step' which can be supplied boundaries for a model, and then searches for an optimal model by adding and removing parameters. This selection occurs through the Akaike information criterion, which is a value that indicates better fit as it grows smaller, allowing it to be used to compare models. Step generates different models, and keeps the one with the smallest AIC. The end result is an improved model compared to the model that was entered.

There are some shortcomings to 'step'. For example, the result for starting with a small model and adding parameters, and the result for starting with a large model and removing parameters are not necessarily the same. This suggests that the algorithm encounters local optima, which it is unable to pass. However, by using 'step' the computer can do the model selection.

After using 'step' to select a model, producing a model that is at least locally optimal, the degree of explanation can be determined via the adjusted R^2 measure of fit, which is a percentage indicating the amount of variance explained. Additionally, the p-values of the individual parameters can be observed to see how significant each parameter is, and the values of each parameter can be observed, although since the technique calculated the parameters based on the data, there is no conscious decision for model parameters, and therefore no underlying reason, making parameter interpretation questionable.

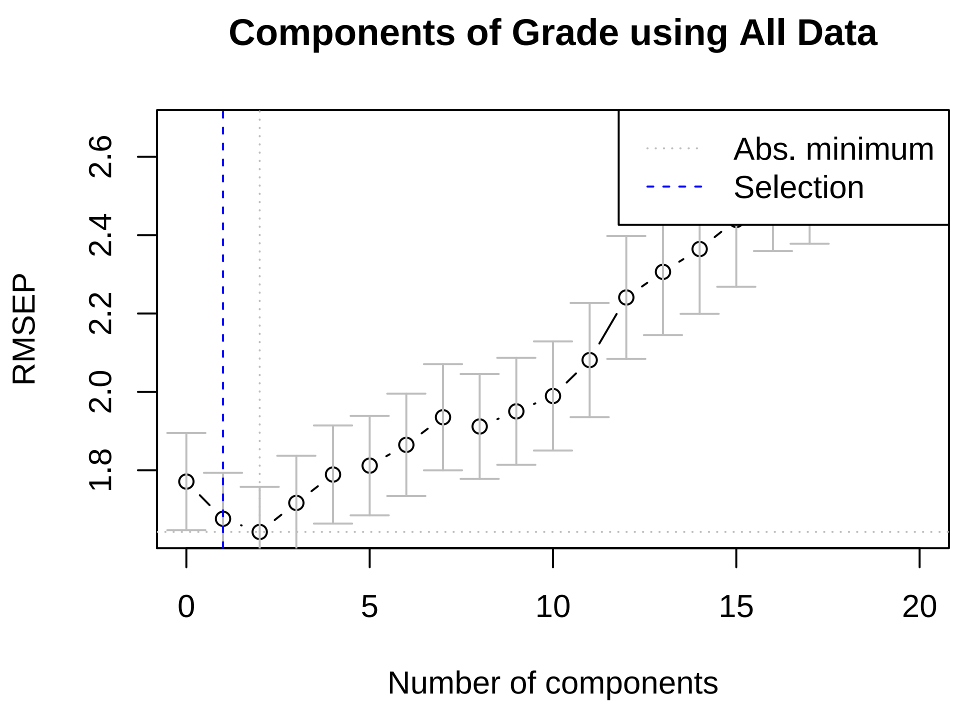
1.9.2 Partial Least Squares Regression

Partial Least Squares Regression is in a sense just like linear regression. However, unlike linear regression, it decomposes the model variables into components. These components are a linear combination of variables that form an orthogonal basis in the data. These components then receive a parameter similar to a linear model, and they constitute the eventual model.

The way these components are chosen is to maximize the amount of information they explain in the resultant variable. The first component explains the most, and subsequent components are used to explain any remaining variance. This way, the result is not a model fit in the same way that linear regression provides, but rather the best possible model given the provided data.

However, creating these components causes an issue: overdetermination. By letting the variance in the resultant variable determine the model parameters the model will eventually fit the data exactly, but when that is the case, there is no information in the model left. To combat that, it is common to cross-validate each model, and use that information to select the number of components to use for a sensible model.

In cross validation, the model is recalculated while leaving one or more data points out. In this case leave-one-out cross validation is used, which recalculates and compares the models once for each data point, leaving that one data point out. The result is an accuracy score for each component, which can then be displayed in a scree plot.



This is such a scree plot, which using the cross validation score (on the y-axis) searches using a heuristic for the point where the cross validation score is approximately the lowest. Here the one sigma heuristic is chosen, which selects the first component within one standard deviation of the optimal model with the lowest cross validation score. That number then becomes the number of components that should be used for a sensible model.

Using the information from this component selection heuristic, it is possible to select a number of components, for which the technique then presents the amount of variance in y that was explained (grades in this case), and how much variance in x was used to explain that (metrics). In this way it gives an indication of how good a model can be, and provides another value for the possible amount of explained variance as a second opinion to the results from the linear model.

1.10 Metric Collection

This thesis attempts to use object oriented metrics in various research questions. However, for those questions to be answered, there must first be metrics to calculate with. Because a self-created metric calculator is a sizeable development effort, which includes a large testing effort to ensure accurate results, the choice was made to search for an existing metrics calculation tool.

The first direction for metrics calculation tools was via the term 'Static Code Analysis'. This area includes tools such as PMD (PMD, 2019), Infer (Facebook, Inc., 2019), SpotBugs (SpotBugs, 2019), and Checkstyle (Checkstyle, 2020). However, these programs do not calculate metrics (or more accurately, this thesis did not succeed in making these tools produce metrics).

All of the above programs are programs that look at either source code or byte code, and match often-occurring patterns of coding mistakes (such as dereferencing a field that might be null) or style errors (like getter names) and report those when they are run. This makes these tools effective at preventing those kinds of errors from reaching production, allowing program maintenance effort to focus more on true logic errors.

Although such information is very useful in various applications, for the goal of calculating object oriented metrics, they are less so. For calculating object oriented metrics, two other tools were found. Designite (Designite, 2019) and JaSoME. (Rod Hilton, 2019) Both of these tools are able to work with Java code (which is used in the course the source code is taken from), and both support a number of important metrics, among which, notably, the Chidamber & Kemerer metrics suite.

Ultimately, tests revealed that Designite was the program most capable of dealing with student code, meaning that with Designite more data would be available. For that reason, Designite was chosen to be the metrics generation tool for this thesis. However, that did mean that not all metrics found in theory could be applied, as Designite at the time of use was only able to calculate the following metrics:

Project Level Metrics	
MCOUNT	Number of Methods
TCOUNT	Number of Types/Classes
Method Level Metrics	
MLOC	Method Lines of Code
CC	Cyclomatic Complexity
PC	Parameter Count
Type Level Metrics	
NOF	Number of Fields
NOPF	Number of Private Fields
NOM	Number of Methods
NOPM	Number of Private Methods
TLOC	Type Lines of Code
WMC	Weighted Methods per Class
NC	Number of Children
DIT	Depth of Inheritance Tree
LCOM	Lack of Cohesion of Methods
FANIN	Number of types using this type
FANOUT	Number of types used by this type

These metrics provide most of the C&K metrics, except for response for a class. FANIN and FANOUT, being respectively the amount of import statements, and the amount of times a class is imported by another class, somewhat cover this area. Number of (private) fields and number of (private) methods are included, which are used in the MOOD metrics suite, but the MOOD metrics themselves are not calculated.

Furthermore, each method gets a cyclomatic complexity, a measure for the nesting depth of if-statements, and the number of parameters passed to a method. Each type and method receives its lines of code, and each project receives the number of methods and types it has. Although some of these metrics were covered in the theoretical framework, the following methods were not, and are given an interpretation on what information they could provide:

Number of Methods The amount of methods used on its own may not be very informative, but projects using fewer methods for the same functionality are either more efficient in their interpretation, or more likely, less able to separate functionality.

Number of Types Similar to number of methods, using more types to implement the same functionality shows a greater ability to distribute responsibilities among pieces of code, though it could also indicate a more diffuse understanding of the problem the code is meant to solve).

Cyclomatic Complexity Cyclomatic Complexity is an indicator for the depth of nesting and the amount of if statements a method has. Deeper nesting increases score more than multiple if statements, in general. It also indicates either a lack of subdivision, or a more confusing piece of code. In general high cyclomatic complexity is considered negative, as it makes the code harder to understand and maintain.

Parameter Count The number of parameters passed to a function. Having more parameters means that more functionality is concentrated in one place, and implies increased complexity. Having many parameters, especially in Java where the class itself is an implicit parameter, is considered a negative thing.

Number of Fields Having a large number of fields means that a lot of data is necessary to make a class work. Each of those pieces of data interacts, and each method in that class is able to change that data. More fields imply more complexity and increased maintenance effort.

Number of Private Fields Number of private fields is the number of fields that can only be accessed from inside the class. Having more private fields relative to the total number of fields is seen as a positive, as it means the state of those fields is managed within the class, allowing the class itself to maintain its own conditions about the state, yet having many fields on its own is a negative. Because of this, there are multiple interpretations possible, meaning that this metric must be analysed in context.

Number of Methods Having many methods means that the class has many different functions, and may indicate that responsibilities are not sufficiently divided, but since this metric does not exclude, for example, getters and setters, two methods that appear for every field, and because it ignores the private methods, which are meant for internal operations only, this metric too must be interpreted in context.

Number of Private Methods Using a large amount of private methods means that code is also subdivided in classes, indicating good organization of source code even at the method level. A larger amount of private methods can be seen as a positive.

Fan-In The amount of other classes from the analysed system this class uses. This is an indicator of coupling. High coupling makes it difficult to reuse a class, which is one of the promises of the object oriented paradigm, and therefore considered a negative. It is possible to argue, however, that a program will rarely share business logic between different programs, and that library classes are often designed for independence regardless. Either way, a high Fan-in means the class can not be reused without also using the classes it imports, which is treated as a negative.

Fan-Out The amount of other classes from the analysed system this class is imported by. This metric shows that a class is depended on by many other classes, meaning that it holds a central role in its logic and its functionality. It also means the class is reused often, while at the same time posing a maintenance risk, as when it is changed, many other sections of code could break. In general, however, reuse is considered a positive, therefore so is a high Fan-out.

These metrics, together with the ones listed above, for which the description can be found in the theoretical framework, will serve as the metric data by which this thesis attempts to answer its research questions. While they have been given a description about what their value might mean, those descriptions are interpretations based on more general ideas in object oriented programming.

It should be noted that the theoretical framework did not discover an experimental foundation for the above descriptions. Rather, they attempt to capture the intent the authors of the metric might have had when creating the metric, and attempts to give a direction towards how one might go about predicting the way these metrics could behave when they are used in an experiment.

The metrics listed above are the ones that were generated for this thesis, and they are what the analyses in this thesis are be based on. Some metrics that would be desirable to use based on theory are missing, whereas some metrics that are present do not have a prominent role in theory. While it will not benefit this thesis, perhaps future research will be able to stay closer to theory with the metrics it uses.

1.11 Hypotheses

This thesis attempts to answer the following questions:

Is there an effect of groups on project grades?

Is there an effect of groups on project metrics?

Is there an effect of teaching assistants on project grades?

Is there an effect of teaching assistants on project metrics?

Is there an effect of assignments on project grades?

Is there an effect of assignments on project metrics?

Is it possible to explain grades by means of metrics?

In this section, hypotheses to these questions are presented.

1.11.1 Is there an effect of groups on project grades?

What this question asks is if an effect will be found of what groups do on their grades. The grade systems here are between 0 and 10 and between 0 and 1, inclusive, which indicate a degree of completion. Provided that students do not copy each others entries line for line, the resulting degree of completion should be at least somewhat different (and if they do they're likely to be removed from the course, thereby removing them from the results). Hence, the expectation is that indeed there will be a difference between grades groups receive, and that there will be an effect of groups on project grades.

1.11.2 Is there an effect of groups on project metrics?

In line with the argument made for the effect of groups on project grades, the effect of groups on project metrics is dependent on the degree of difference students apply to their projects. Importantly, these differences go beyond the level of requirement compliance. Because there are multiple designs that can fulfil a certain requirement, and because the requirements do not call for the 'minimal' implementation, if that is even possible, the expectation is that there is not just variation with respect to requirements, but that there is variation in projects in general, which is why the expectation is that indeed there will be an effect of groups on project metrics.

1.11.3 Is there an effect of teaching assistants on project grades?

Although it would be a valid interpretation, this question does not ask if the existence of teaching assistants changes the grade. These projects would not have grades without teaching assistants. Instead, it attempts to see if their grading was uniform - if there were no differences between what grades they gave in the various categories (OO standard, coding standard, and (corrected) grade).

As stated, it is almost a judge of character: can these teaching assistants be trusted to grade consistently. However, the circumstances of grading are not a test of character, and not an exercise in consistency. The grading conditions call for grading a sizeable amount of projects within 8 hours, which is why there is little communication, meaning reduced opportunity for reducing personal bias. Because of this, it is expected that some form of bias can be found in the collected data, hence, that there is an effect of teaching assistants on project grades.

1.11.4 Is there an effect of teaching assistants on project metrics?

In research it is rarely ideal to outright reject possibilities, which is why this section discusses the possibility that for certain teaching assistants, students asked a large amount of questions, and as a consequence these teaching assistants had a role in their final products large enough that such an effect might be found. However, given that none of the coding is performed by teaching assistants, that suggestion seems unrealistic. Therefore, the hypothesis for this question is that there is no effect of teaching assistants on project metrics.

1.11.5 Is there an effect of assignments on project grades?

There are arguments for both the presence and the absence of an effect here, such as the incremental difficulty in the assignments that would speak for consistent grades, but also the unknown factor of student progress. Because this question depends greatly on the way students make the assignments, and because that sort of analysis is not part of the theoretical framework, the safe hypothesis is to say that there is no effect of assignments on project grades.

1.11.6 Is there an effect of assignments on project metrics?

For the assignments, the idea for the first assignment, RPG, was to introduce inheritance and give students a hands on experience with using it in a program. This means that they are pushed to use inheritance. Then, for Card Game, there is no such push, and instead they are asked to take a GUI implementation and base their programs of that. Finally, they do Graph Editor, which both asks for making a GUI, and for making an own project design. This means that gradually, the metrics should converge to programmer preference, and that there should be an effect of assignments on project metrics.

1.11.7 Is it possible to explain grades by means of metrics?

Object oriented metrics were conceived to indicate certain desirable qualities in program design for specific purposes. However, for the course, students are learning about possible desirable qualities, and the assignments do not call for their use - only their awareness, meaning that the resulting programs do not necessarily use them. As a result, the given grades should not reflect these metrics, and hence it should not be possible to explain grades by means of metrics.

1.11.8 Hypothesis summary

In this section the research questions for this thesis were given a hypothesis. Summarized, these hypotheses are the following:

Is there an effect of groups on project grades?

There should be an effect of groups on project grades.

Is there an effect of groups on project metrics?

There should be an effect of groups on project metrics.

Is there an effect of teaching assistants on project grades?

There should be an effect of teaching assistants on project grades.

Is there an effect of teaching assistants on project metrics?

There should not be any effect of teaching assistants on project metrics.

Is there an effect of assignments on project grades?

There should not be an effect of assignments on project grades.

Is there an effect of assignments on project metrics?

There should be an effect of assignments on project metrics.

Is it possible to explain grades by means of metrics?

It should not be possible to explain grades by means of metrics.

Chapter 2

Factor Analysis

This chapter attempts to answer the following research questions:

Is there an effect of groups on project grades?

Is there an effect of groups on project metrics?

Is there an effect of teaching assistants on project grades?

Is there an effect of teaching assistants on project metrics?

Is there an effect of assignments on project grades?

Is there an effect of assignments on project metrics?

To do so, this chapter looks at the results of performing ANOVA using the categories of (Student) Group, Project (Name), and Teaching Assistants (categorized as A, B and C), and analyses if there is an effect of these categories on both accumulated metric data and grade data. Then, it tries to determine the size of each effect using Tukeys Honest Significant Differences.

2.1 Meta-Analysis of ANOVA

Because of the number of ANOVAs that needed to be performed, they were performed via R script. The results thereof can be found in the table with results for all data and the table with results for groups that finished all assignments. The following summary of results shows, given which assumptions one is willing to relax, the amount of significant ($p < 0.05$) results that are available.

This table has 2 sections, one for all data and one for complete cases, in which, for each category tested, and for each reported level of met ANOVA assumptions, the number of significant ($p < 0.05$) ANOVAs relative to the total number of ANOVAs, as well as the percentage of significant ANOVAs.

	All Assumptions		Homogeneous, not Normal		No Assumptions	
	ratio	percentage	ratio	percentage	ratio	percentage
All Data						
Group	9/189	0.048	48/189	0.254	56/189	0.296
Project	1/189	0.005	16/189	0.085	115/189	0.608
TA Team A	0/189	0	7/189	0.037	16/189	0.085
TA Team B	1/189	0.005	3/189	0.016	14/189	0.074
TA Team C	1/189	0.005	11/189	0.058	26/189	0.138
Complete Cases						
Group	12/188	0.064	38/188	0.202	41/188	0.218
Project	1/188	0.005	23/188	0.122	116/188	0.617
TA Team A	0/188	0	7/188	0.037	20/188	0.106
TA Team B	2/188	0.011	24/188	0.128	35/188	0.186
TA Team C	2/188	0.011	29/188	0.154	42/188	0.223

These results clearly show that there are problems with the ANOVA-approach. Almost none of the models made satisfy the tested assumptions. For those that do the percentage of models that are significant is often smaller than 0.05. Since a p-value indicates the likelihood that such an observation happened by chance, at a threshold of 0.05 there is 5% chance that a model is found by chance. Therefore, any data that produces fewer than 5% significant models should be treated as gathered by chance.

When looking at the table of ANOVAs with all assumptions met, fewer than 5% of models are significant. Many models are 'homogeneous, but not normal'. Each category has at least one categorization that produces enough results to be considered valid. However, for the increased amount of significant results, the validity decreases with violated assumptions.

It must be said, however, that the tests to find these assumption violations are not well suited at dealing with large sample sizes, because those cause them to detect violations with a very small effect. And so these 'tests' actually don't work very well in an automated environment. This effect is worse with larger sample sizes.

For this thesis, the sample sizes are around 60 for Project, 30 for TA categories and 3 for Group. This means that Project and TA are adequate, making it possible that the tests are too sensitive for them, whereas with Group there are probably real issues. This means that for Project and the TA categorizations it's probably fine to use the results where both assumptions are violated, whereas for groups that is likely a bad idea.

In conclusion, all factors in the data have an effect on said data. Because of the nature of ANOVA, the size of this effect is unknown at this time. Further investigation is necessary. The presence of the effect is measured by relaxing assumptions, reducing its predictive strength. However, because the way data was gathered pulls such predictions into question regardless, these results should be treated as angles of approach to future research, which means that there is still merit to continuing to look for effects.

2.1.1 Results of ANOVA

Based on the meta analysis there are some remarks that can be made about the validity of the results. In particular the way the assumptions are violated makes these results less reliable. However, because these results will likely not be representative for other situations, and because there is still a sufficient amount of results, it is still worth it to look at what the results are.

Below, for each general category (Groups, Projects and Teaching Assistants) the results are summarized. In there, the individual data columns that showed as significant in ANOVA are discussed, along with why that particular piece of data might show up as significant in the context of the used category.

2.1.2 Group Results

Of all the data, groups with their 81 different possibilities have the highest level of variability of any category, and also the fewest amount of data points. Despite this small sample size, the results are relatively able to meet assumptions, and even produce a decent amount of significant results, which is what will be discussed in this section.

For this section there are some data columns that are significant for both complete cases and all data, and some that are only relevant for one of those data sets. Because of this, there are two angles that can be explored using these results: The general way groups did these assignments, and the difference between groups that completed the course, and those that didn't.

Similarities in Group Data

This section discusses the things that groups had in common, which means that it describes the general way in which students worked on projects. Since these results are common between both groups that completed the course and groups that didn't, and since most groups did complete the course, this data is somewhat skewed towards the students that completed the course, and may only represent the groups that did complete the course. Either way it is a good indicator of what this data did among groups that took part in the course.

As per usual, grades differ between groups: OO Standard, Grade and Corrected Grade all register as significant. Coding Standard was not significantly different, implying that in terms of their basic code quality indicators such as variable names and indentation groups weren't very different in this version of the course. On the other hand, their capabilities with object orientation and their ability to produce a correctly functioning program did differ from group to group.

In terms of metrics, some groups used many (nested) if statements per unit of code as seen from the high cyclomatic complexity found in some pieces of code. The sizes of both types and methods also had significant variability, as seen in the Method- and Type Lines Of Code. If this was reflected in grades, it was in the OO Standard, as that also differed between groups, but this is typically a Coding Standard aspect, which makes it interesting to see what metrics (seemingly) contributed to that grade.

Speaking of OO Standard: various metrics that would be related to that grade show that the amount of (private) fields and (private) methods (NOF, NOPF, NOM, NOPM) varies greatly between groups, where larger numbers on either of these are typically an indicator of too many responsibilities of a class. Here too, it would be interesting to see if these significantly contribute to the OO standard.

Other metrics that are interesting in the context of OO Standard are Weighted Methods per Class and FANOUT. Both of these metrics varied greatly between groups. Some had high WMC, some low WMC. Some had high fan out, some had low fan out. Since low WMC is considered a positive, and high fan out an indicator of good design at the expense of a larger maintenance effort, these metrics too would be expected to contribute to the OO Standard score.

In summary, the grades differed between groups, which is expected. Furthermore, cyclomatic complexity and lines of code, metrics that would be expected to affect the coding standard score did apparently not do so, despite significant variability in these values between groups. Vice versa, significant variability in OO Standard was found, as well as variability in NOF, NOPF, NOM, NOPM, WMC and FANOUT, which going from the theory would implicate a connection, though at this time that is not more than an assumption.

What does show from these results is that there are significant differences in the way students did their projects, showing very different results across a large number of metrics, and working by the grades, leading to very different results.

Differences in Group Data

Some results in analysing groups were found in one data set, but not in the other. This section will be looking at which results those are, and what they might indicate about the difference between students that completed the course, and students that did not. Strictly speaking, this is not the correct analysis for this kind of result, because there are comparative techniques that are more appropriate for comparing groups, but there are nevertheless different results.

The results for all data show that in all data the large classes are more centralized in a few groups, from the significant difference in both maximum and Alves very high risk on Type Lines Of Code. Although this is somewhat congruent with the differences in sizes, because it concerns both these ways of counting the results, the larger outliers are more prevalent in all data, than they are in complete cases. Since larger classes tend to be less well designed, this might be an indicator that these students did not grasp the material well enough, were, as a result, unable to complete the assignments, and left the course.

Another difference is in depth of inheritance tree. When using all data, it only shows that some groups had more slightly larger inheritance trees, from looking at Alves Medium Risk, which differs significantly. Large and small trees seem distributed about evenly, but medium trees are not. On the other hand, for complete data it is Alves Low Risk that registers as significant, indicating that the larger trees are evenly distributed. Since larger trees are treated as good design, this result seems to contradict the previous explanation that students did not grasp the material.

The final comparison for data sets comes from Fan-In. Both data sets have significant results here, but where all data shows low, medium and high risk for Alves rankings as significant, complete cases shows median fan-in as significant. In other words, all data has a similar amount of imports across groups when many imports are used, but when that number shrinks the imports start to vary.

For complete cases, this effect does not show. The only thing that shows is that the median is different for some groups (but the mean does not show as significantly different), which indicates that the distribution of imports changes, where some groups may have a continuous, a linear, or an exponential distribution of imports.

These points are not necessarily incompatible, but because they show in different manners there is a difference here, where all data seems to show more clearly that the amount of import statements differed compared to using only complete data. In general, this too is a negative indicator, as many imports makes a system more dependent on other classes, increasing maintenance effort.

In summary, there are some differences in the results from both data sets, but those results only vaguely point to differences between the data sets, and do not seem to indicate very clearly why these data sets are different. This does not mean that those reasons can not be found, as the used analysis was not chosen to find said differences, but these results do not seem to indicate them.

Conclusion

This section is related to the following research questions: Is there an effect of groups on project grades?
Is there an effect of groups on project metrics?
To both of those questions, the answer can be given that, yes, there is an effect. 3 of 4 grades and most recorded metrics showed a significant difference between groups.

At this stage, it is not possible to quantify the size of these effects. Moreover, despite attempts to provide an explanation for the results, the given explanations seem to contradict, and should therefore probably be ignored. What's more, given the sample size of the ANOVA, there are concerns about the validity of these results. Concerns aside, not only were differences found, they also showed that there were many different approaches to the way students did this course, producing different results that received different grades.

2.1.3 Project Results

Projects, having the fewest amount of different categories, was positioned to generate many significant results with ANOVA, and with about 60 % of the analyses significant, that prediction has come true. Indeed, there are so many differences that they can be reported by category, rather than by a specific aggregator in a category, which is how these results are reported. Moreover, these results apply to both data sets (complete cases and all data).

As far as grade is concerned, only coding standard shows as significant, which means that grades were fairly consistent across assignments. The fact that coding standard is different is an interesting result because graph editor does not record coding standard, meaning that there is a difference, and it exists between RPG and Card Game. This difference will be explored further in the results for TukeyHSD.

In terms of metrics, the projects differed in the amount of methods and types they had. Since they are intended to have different functionality, this result is expected. For this category too, a look at the TukeyHSD results ought to show in what way these metrics changed between projects to give an indication of the relative complexity of each assignment.

The same complexity analysis can be made for Method Lines Of Code, Cyclomatic Complexity, Parameter Count, Number Of Fields, Number Of Methods and Number Of Private Methods. All of these values are somewhat indicative of the direct complexity of classes, which is why further research is needed to reveal which was the more complex assignment.

Another area where Number Of Fields, Number Of Methods and Number Of Private Methods can be used, along with Number Of Children, Depth of Inheritance Tree, Lack Of Cohesion Of Methods, Fan-In and Fan-Out, is in how the design quality changed between projects. As these metrics get higher, the design is considered to be worse, which is an alternative supporting argument for the complexity of assignments.

Although it is necessary to look at these results in more depth, which will happen in the section on TukeyHSD, with this data it is already possible to generally answer the relevant research questions for this section: Is there an effect of assignments on project grades?
Is there an effect of assignments on project metrics?
The answer for grades is a partial yes, there is an effect (for coding standard). The answer for metrics is a larger yes, for 17 of the used metrics, on many different methods of aggregation. This section cautiously produces a double confirming answer of its research questions, which will be fleshed out with details later on.

2.1.4 Teaching Assistants Results

For Teaching Assistants (using either data set), regardless of the way of counting, this analysis shows that they give different scores for OO Standard and Coding Standard. Now, because these teaching assistants do not grade the same assignment that is potentially because their groups indeed had different quality code. Still, this is a result that could have been expected, and has been found.

What has not been found: A difference between grades. Despite their quality standards differing, teaching assistant grades were found to be statistically similar. Moreover, most metrics that seem to correlate to teaching assistants do not do so across all countings (A, B and C), and not across both data sets. There are two exceptions: Number of Types and Number Of Private Fields.

Some potential explanations are: Some groups that used fewer, or more classes or fields were disproportionally assigned to certain groups of TA's. TA feedback suggested using more or fewer classes, or using more or fewer fields. Either way, this is something that would not be expected, and warrants further investigation.

Metrics

Significant differences between teaching assistants occur on seemingly arbitrary aggregations of metrics, such as number of methods, minimum method lines of code, number of fields percentile medium risk, and so on. What these seem to indicate is that the projects different teaching assistants were assigned to differed significantly, but only on some specific points. The most likely explanation seems to be accidental patterns from the way projects were assigned.

Given that the differences between projects are very extensive, and that some teaching assistant groups only graded certain specific projects, the hypothesis that the found differences were caused from the way projects were distributed is strengthened even more. Because of this looking only at these ANOVA results where the relevant categories are ignored is not valid.

To ensure that the results are evaluated in a sensible manner, they must be observed individually, which will happen later with TukeyHSD. This section can only conclude that there are significant effects from teaching assistants on metrics, but it can not determine if those events are systemic errors at this stage.

Conclusion

In this section, answers were sought for the following research questions: Is there an effect of teaching assistants on project grades?

Is there an effect of teaching assistants on project metrics?
To the question of the effect of teaching assistants on grades, the answer is that indeed, an effect exists. Teaching assistants gave different grades for OO Standard and Coding Standard. However, no significant difference in the final grades teaching assistant was found.

To the question of the effect of teaching assistants on metrics, the answer can not yet be given, as there is a plausible explanation that assigns the found results to a systemic error that the ANOVA data alone is unable to analyse. For the answer to that question, further investigation with TukeyHSD is necessary.

2.1.5 Conclusion of ANOVA

In this section, ANOVA was used to look at the effects different groups have on the values of metrics and grades. This concerned the following research questions.

Is there an effect of groups on project grades?

Is there an effect of groups on project metrics?

Is there an effect of teaching assistants on project grades?

Is there an effect of teaching assistants on project metrics?

Is there an effect of assignments on project grades?

Is there an effect of assignments on project metrics?

For each of these questions it was shown that yes, there is some measured effect. In the case of grades it was not a complete yes, as there was no category that showed an effect in all categories, and the same is true for metrics, where only a few metrics showed significant. Nevertheless, for all of these questions there was some effect.

Interesting observations are that these results seem to show that students were able to organize their own code, that the projects required different approaches, and that there is a difference in the grades student assistants give. While this is already a usable result, this analysis does not concern the size of the result, which is another angle that can be evaluated.

2.2 Tukey Honest Significant Differences

Because ANOVA is not a test for the size of an effect that a category has, to determine the size of the effect, TukeyHSD is performed. To get an impression of the number of differences involved, the first procedure is to generate an overview of all TukeyHSD tests. This is accomplished by calculating TukeyHSD for every significant ANOVA and counting the number of significant differences reported.

In the following table, for both all data and complete cases, subdivided by the number of assumptions of ANOVA met, the number of tests for significant differences, the number of differences that were reported significant ($p < 0.05$), and the percentage that ratio correspond to are listed for each category tested.

	All Assumptions		Homogeneous, not Normal		No Assumptions	
	ratio	percentage	ratio	percentage	ratio	percentage
All Data						
Group	601/181440	0.003	1960/181440	0.011	2567/181440	0.014
Project	2/343	0.006	31/343	0.09	240/343	0.7
TA Team A	0/240	0	7/240	0.029	29/240	0.121
TA Team B	0/210	0	0/210	0	26/210	0.124
TA Team C	1/894	0.001	36/894	0.04	72/894	0.081
Complete Cases						
Group	274/48216	0.006	674/48216	0.014	770/48216	0.016
Project	1/346	0.003	37/346	0.107	225/346	0.65
TA Team A	0/300	0	11/300	0.037	30/300	0.1
TA Team B	0/525	0	47/525	0.09	83/525	0.158
TA Team C	2/1470	0.001	64/1470	0.044	105/1470	0.071

If it was not clear before what it means to look at the differences between pairs of categories, it must be now. The amount of comparisons grows by N^2 with the amount of categories, so with 81 different groups the amount of comparisons is in the order of 100000.

No matter how many assumptions one is willing to violate, the Group category does not reach 0.05% significant differences, and so it's likely that any significant differences are indeed by chance. Therefore, it is not helpful to look at the individual differences for groups, despite the promising results from doing ANOVA on them. Therefore, from here on out the effects from groups will be ignored.

Projects on the other hand will not be ignored. There are plenty of significant differences between different projects. In fact, if one is willing to ignore most assumptions (and given the sample size that is a valid decision), as much as 70% of all differences is significant. That means that the intended differences between projects not only exist, they are also plentiful.

Finally, there are some curious results for the TA categories. Firstly, they manage about 10% significant differences. Relaxing all assumptions here is not necessarily valid, but when they are relaxed, they fairly comfortably hit the threshold. There seem to be no major differences between categorizations, with the way categorization C works making it the preferred perspective.

2.3 Results

After all this there are still about 500 differences to observe. These 500 results are not just metrics with a certain grouping, but also only apply to the difference between two exact groups. The result is a sparse matrix with a difference here and there, which can be compressed to a more dense representation, but that produces just a list of seemingly unrelated differences, which may come across as confusing.

The full results may be browsed in the tables, which for future research should provide angles of research. However, here only some of the more interesting results will be mentioned. The full results: 4.1, 4.2, 4.5, 4.6 4.9, 4.4, 4.7, 4.8, 4.10, 4.3.

2.3.1 Project Results

There is one indication that students improved over time: Their Coding Standard grades increased from RPG to CardGame (though there is no data for GraphEditor). There are also a number of other indicators that the assignments got harder:

- Total Number of Types (Class files) is on average 6.5 higher than RPG for cardGame and 11 higher than RPG for GraphEditor. More types means more responsibilities and therefore a more complex program.
- Total Number of Methods is on average 40 higher than RPG for cardGame, and 66 higher than RPG for graphEditor. More methods is also an indicator of increased functionality and therefore complexity.
- The amount of low risk classes for Method Lines of Code, according to the Alves-method, decreased very slightly with consecutive assignments, meaning that a larger portion of methods was larger, implying that each method was on its own more complex as well.
- The maximum amount of methods increased in consecutive assignments, which means more implemented functionality.
- The standard deviation for number of methods increased in consecutive assignments, meaning that there was more variety in class complexity. Some types exclusively for product types, some more complex data types. Complexity starts concentrating in certain places.
- The standard deviation in number of private methods also increases in consecutive assignments for less even distribution of complexity.
- The minimum type lines of code increases with consecutive assignments, implying that the number of features in a class increases.

Besides the seemingly more difficult assignments, there are some indicators that imply that both assignments got harder or that students improved, or both:

- Cyclomatic complexity did increase after rpg, again implying more complexity. Interestingly, Graph Editor is equally complex in terms of the number of branches as Card Game. Since it is in a sense more complex, it seems that students get slightly better at dealing with the complexity.
- The maximum and mean Lack of Cohesion of Methods are the lower for CardGame than for the other assignments. This indicates that the single responsibility principle was more straightforward to apply for CardGame. Coincidentally, it was the hardest for RPG, which with it being the first assignment should not be surprising. The reason for the regression during GraphEditor could be attributed to the difficulty of that assignment.

There are also some observations that imply that CardGame was harder than GraphEditor. This is not entirely surprising as both assignment produce a similar end product, but it's not in the line of expectations. These observations are the following:

- The maximum Parameter Count is highest for CardGame, after GraphEditor and RPG. Fewer parameters generally means lower complexity, but that is not consistent with the complexity progression from other results. A plausible explanation is that the problem requires more parameters.
- The FAN-IN and FAN-OUT tend to be largest for CardGame, after GraphEditor. As a metric for coupling that makes CardGame the more complex exercise.

- The mean Method Lines of Code are highest for CardGame, after GraphEditor and RPG. This again points to CardGame being more complex (or students improving after CardGame)
- The Percentile Low risk for number of fields is lowest for RPG, highest for CardGame, meaning that classes in CardGame carry more complexity.
- The maximum number of (private) methods is highest for CardGame, after RPG and GraphEditor, showing that in terms of the amount of class functions CardGame is the most complex.

The safe conclusion is to say that RPG is the easier assignment, and leave the conclusion for CardGame open. For those same reasons claiming that students improved is also questionable.

Finally there's a result about the importance of inheritance, as metrics that indicate its use decrease sharply after RPG.

- The Number of Children seems to decrease from RPG, and then stay about the same. RPG recommends inheritance heavily, but inheritance is not really that common in practice.
- The maximum Depth of Inheritance Tree decreases with consecutive assignments, strengthening the suggestion that later assignments need less inheritance.

It is by design that RPG uses inheritance extensively, and these results confirm the effectiveness of that approach.

In general these results seem to show that some of the improvement that is intended to take place during this course does take place. Some of the design goals of the assignments also come across, and the hypothesis of the course goals seems to sufficiently explain these results. The one exception is that some results show GraphEditor as the harder assignment, whereas some results show CardGame as harder. Because of this, it might be sensible to re-evaluate the design of these assignments.

2.3.2 Teaching Assistants

For teaching assistants, the expected differences should be from the grades they gave. Although they are able to influence end products through answering questions, they are not the ones making said products, which means that there should not be any effect on the metrics calculated from those end products. And yet, some metrics still register as significant.

Grades

As far as the grades are concerned, TukeyHSD is not able to determine a significant difference for both the grade and corrected grade from different groups of teaching assistants. It is, however, able to determine a significant difference for some groups of teaching assistants between the OO Standard and the Coding Standard they assign.

Because of how the groups of teaching assistants were organized, categorization C actually does not have data on OO Standard and Coding Standard, which is because the combinations of groups found in categorization C only appear for graph editor - an assignment for which neither OO Standard nor Coding Standard grades were assigned.

The most significant results are found by looking at all data, where regardless of categorization A or B, the results are the same (as without graph editor they are indeed identical). On the grade for OO Standard:

- A grades higher than C, and E
- D grades higher than B, and C, and E

On the grade for Coding Standard:

- A grades higher than C, and D
- E grades higher than C, and D
- F grades higher than C

In contrast, complete data does not have this amount of significant differences. Complete data produces just one result: A grades higher than C on OO Standard and Coding Standard - something that was also visible from using all data.

Because of the way OO Standard and Coding Standard were gathered, this data is only relevant for the first two assignments: RPG and Card Game. Because using all data means results from students that left after either of those assignments is included, and because students that left the course likely received lower grades, having more differences when using all data is an expected result.

The consequence is that a conclusion that takes all those results in account can only conclude that students that were graded by TA Team C got lower OO Standard and Coding Standard grades than students that were graded by TA team A. This is consistent, but it is also skewed towards the students that completed the course. This is the consequence of having missing cases.

As for the reason these differences appear, there can be multiple reasons. Perhaps teaching assistants were stricter, and gave lower grades for that reason. Perhaps teaching assistants were assigned groups that just made more errors, meaning that they did not necessarily grade more strictly, their groups just did not perform as well as other groups. However, in that case they should also have given significantly lower grades, but grades did not show up as a significant difference. Perhaps there is another explanation, but it seems that TA team C was indeed simply stricter than TA team A.

Metrics

When discussing the results for teaching assistants there was little mention of the results for metrics. Although there were significant differences between groups of teaching assistants, they occurred on seemingly arbitrary aggregations of metrics, such as number of methods, minimum method lines of code, number of fields percentile medium risk, and so on. This section tries to see what the magnitude of those results is, and tries to explain why they occur in more detail.

There is a rather large amount of differences that can be discussed in this section. Certainly, these results were found, they are significant, and therefore valid, however, they are not very capable of explaining why these differences occurred, and that is what this section attempts to do. Because there are new developments here, it is useful to create a separate hypothesis for these metrics to explain why the found differences occur.

In this case, it would be safe to assume that because not all groups had the same grading load, their metrics differ due to differing sample sizes, and that between groups that had similar work load these differences are absent. This way, comparing unevenly sized samples can be avoided, which given the sample size is defensible. Following this assumption, the approach is not to look at all differences, but only at differences between comparable groups.

- For categorization A, that means groups A,B,C and D, as group D receives the data of E and F if they worked together with D, giving group D roughly the same amount of work as groups A,B and C.
- For categorization B that means groups A,B,C and F, because categorization B works like categorization A, except grouping all results with the later member.
- For categorization C that means that groups A,B and C; D,E and F; and G,H and I are coupled, as they have the same work load.

Using this approach, 3 results remain for categorization A using all data that conflict the hypothesis that these results are due to systematic error:

- D has 0.64 higher mean Number Of Private Fields than B
- D has 0.1 more very high Alves risk Number Of Private Fields than B
- C has a 1.6 higher maximum number of children than D

What this indicates is that D graded many classes that had a relatively large amount of private fields compared to the class size, and also in general, that were not directly extended often. The lack of extension need not even be the same class. This effect is likely not causes by the systemic error, which means that based on these results the conclusion is that an effect exists. The arbitrary nature of these differences makes identifying a cause difficult, however.

For categorization A with complete data there are more results than when using all data:

- C has on average 9 more Types than D.
- D has 0.69/0.66/0.67 higher Median Number Of Private Fields than A/B/C respectively.
- D has 0.91/0.98/0.91 higher Mean Number Of Private Fields than A/B/C respectively.
- D has 1.05 more Standard Deviation of Number of Private Fields than B.
- D has a 0.06/0.05 lower Alves Medium Risk on Number Of Private Fields than B/C respectively.
- D has a 0.13/0.14/0.13 higher Alves Very High Risk on Number Of Private Fields than A/B/C, respectively.
- D has a 0.53/0.51 higher minimum Number Of Private Methods than B/C respectively.
- D has a 0.09 higher Medium Percentile Risk on Type Lines Of Code than B
- C has 2.4 more maximum number of children than D
- D has 0.03 lower Alves Very High Risk on Fan-In than B

These results are similar to those from using all data, but now that the complete data is used the differences are larger. Because using complete data means that each category has fewer data points, this result is consistent with a systematic error because of low sample size, which a smaller sample size logically would make worse. Having more results show up as significant is another indicator that these effects are caused by a systematic error.

Discussing the differences for B is quick, as after the attempts to eliminate the systemic error from dividing assignments between student groups, there are no significant results from either categorization B with all data, or categorization B with complete data remaining. This is interesting, as it pulls the results from categorization A into question, and strengthens the assumption that these results are due to a systematic flaw.

The same lack of difference is found between TA groups A,B and C when using categorization C, irrespective of which data set is used. Neither of those groups have differing metrics, which is consistent with categorizations A and B. Because these groups certainly had similar work load, these results are very supportive of the hypothesis that these results stem from systemic error.

Although A,B and C did not seem to have differences, when it comes to the differences between D,E and F, however, a similar pattern emerges as with categorization A: using complete data generates more results. In this case, using all data produces no results, whereas using complete data produces the following results.

- D has a 1.43 higher Minimum Number Of Private Fields than E and F
- D has a 2.00/1.70 higher Mean Number Of Private Fields than E/F, respectively
- D has a 0.29/0.24 higher Alves Very High Risk for Number Of Private Fields than E/F respectively
- D has a 0.10/0.08 higher Alves High Risk for Fan-In than E/F respectively
- D has a 0.89 higher Minimum Number Of Private Methods than F
- E has a 0.06 higher Alves Very High Risk for Fan-In than D

Finally, there are the results for groups H, I and F. It should be noted that there is a difference between which results are significant for both data sets. However, this is because the results for which ANOVAs were significant, since those determine whether or not TukeyHSD is calculated for a results. H, I and F only graded Graph Editor, so any differences between them are shared.

- H has 110 fewer methods than G
- G has 22.5 more types than H
- H has a 0.29/0.4 higher Percentile Medium Risk in Number of Fields than G/I
- G has a minimum Type Lines Of Code 6.35 higher than H
- H has a 0.37 higher Percentile High Risk in Number Of Private Methods than G
- H has a 0.29 higher Percentile High Risk in Fan-In than G
- H has a 0.33 higher Minimum Fan-In than G and I

Between the other groups, the differences were not as clear. Perhaps the sample size managed to hide the differences between the projects that teaching assistants graded to where it no longer showed up as significant. However, when the sample size shrinks it becomes clear that there can indeed be differences between what teaching assistants grade.

The important question here, however, is whether or not those differences can be attributed to teaching assistants. In this, the found results are indicative of systematic error, and therefore the conclusion at this time is that these differences are not caused by teaching assistants. Although effects were discovered, their circumstances suggest that it is valid to ignore them.

2.3.3 Teaching Assistants Conclusion

In this section, the results for teaching assistants were more closely analysed. These results were used to answer the following research questions: Is there an effect of teaching assistants on project grades?
Is there an effect of teaching assistants on project metrics?
In absolute terms of results, the answer to both of these questions is positive: There is an effect of teaching assistants on both grades and metrics. However, that conclusion is not nuanced enough to serve as the end conclusion.

The effect that teaching assistants were expected to have, and indeed shown to have, was an effect on the grades they gave. However, this was only expressed in the OO Standard and Coding Standard grades. In general, their grades were similar, but they graded these standards differently. In particular TA team C was stricter than A, though it is uncertain why that is.

Contrarily, no effect of teaching assistants on metrics was expected, yet such results were common. Because of this, the hypothesis needed to be adjusted from 'no effect' to 'systemic effect': There is an effect, but it is caused by the organization of the data. Approaching the problem in this manner showed results consistent with this hypothesis, even those seemingly conflicting with the hypothesis of systemic error.

Although that approach was able to resolve the differences in the data for metrics, it also raises the issue that this systemic error is also present in the grade data. However, that is not consistent with the data, as the difference in grading exists between group A and group C, which have shown no differences in any of the other ANOVAs.

In conclusion, there is an effect in both metrics and grades, however the effect in metrics is likely caused by systemic artefacts. Therefore, there is no real effect of teaching assistants on metrics. Contrarily, even when assuming that this systemic error exists in the grades data too, the results contradict that hypothesis also, which means that there likely is a real effect of teaching assistants on grades. This is consistent with the hypotheses for these research questions.

2.4 Tukey Conclusion

In this section the ANOVA results for Projects and Teaching Assistants were re-evaluated. For Projects, the relevant research questions are the following: Is there an effect of assignments on project grades?
Is there an effect of assignments on project metrics?
Both these questions have been confirmed: These values change. This section looked at the specific metrics that changed, and the size of the change and attempted to create a more detailed conclusion this way.

The conclusions there were that some of the improvement that is intended to take place during this course does take place. Some of the design goals of the assignments also come across, and the hypothesis of the course goals seems to sufficiently explain these results. The one exception is that some results show GraphEditor as the harder assignment, whereas some results show CardGame as harder. Because of this, it might be sensible to re-evaluate the design of the assignments.

With respect to teaching assistants, this section managed to arrive at a better answer than was given in the ANOVA section by analysing the results through the assumption that any found effects were from the organization of the data. As a result, the relevant research questions: Is there an effect of teaching assistants on project grades?
Is there an effect of teaching assistants on project metrics?
Were answered with a 'Yes, there is an effect of teaching assistants on grades.' and a 'No, there is no real effect of teaching assistants on project metrics, all effects seem sufficiently explained by attributing them to systemic error.'

2.5 Overall Conclusion

This chapter attempted to answer six research questions. Below are, in order the questions, their hypotheses, and the answers that this chapter found. found the following answers:

Is there an effect of groups on project grades?

Hypothesis There should be an effect of groups on project grades.

Result There is an effect of groups on project grades - groups received different grades.

Is there an effect of groups on project metrics?

Hypothesis There should be an effect of groups on project metrics.

Result There is an effect of groups on project metrics - groups showed differences that indicated that they varied their program designs.

Is there an effect of teaching assistants on project grades?

Hypothesis There should be an effect of teaching assistants on project grades.

Result There is an effect of teaching assistants on project grades - although teaching assistants graded mostly consistently, they gave different grades for OO Standard and Coding Standard.

Is there an effect of teaching assistants on project metrics?

Hypothesis There should not be any effect of teaching assistants on project metrics.

Result No, there is no real effect of teaching assistants on project metrics, all effects seem sufficiently explained by attributing them to systemic error.

Is there an effect of assignments on project grades?

Hypothesis There should not be an effect of assignments on project grades.

Result Yes, coding standard was different between assignments. To be precise, it increased between RPG and Card Game. Other grades were not significantly different however.

Is there an effect of assignments on project metrics?

Hypothesis There should be an effect of assignments on project metrics.

Result Yes, a large amount of metrics differed between projects, showing the differences between these assignments.

The end result is mostly consistent with the hypotheses. The only difference between the hypotheses and the results is that there is an effect between assignments on coding standard, which increased from RPG to Card Game. This difference is potentially because of students getting used to the grading system, potentially because of personal improvement, which is why more research is necessary to conclusively answer the question of the effect of assignments on grades.

In general it is a good idea to retry this exact research with different data to see if the same results still hold. This study focused on finding effects in the course, which can be valid in their own right. However, the many ANOVAs that did not meet their assumptions, and the inability to control external influences to the used results means that it is difficult to generalize this result to multiple courses or even education of object orientation in general.

With that said, the fact that these results match the chosen hypotheses as well as they do, means the expectation is that any repeat of this research using data from later years of teaching this course, or perhaps other courses, will find the same conclusions, which would imply that these courses are working as intended: teaching students about object orientation, giving the freedom to design ones own programs, and receiving fair grades for them. At least this course seems to have accomplished that.

Chapter 3

Predicting Grades

In this chapter an attempt will be made to answer the following research question:

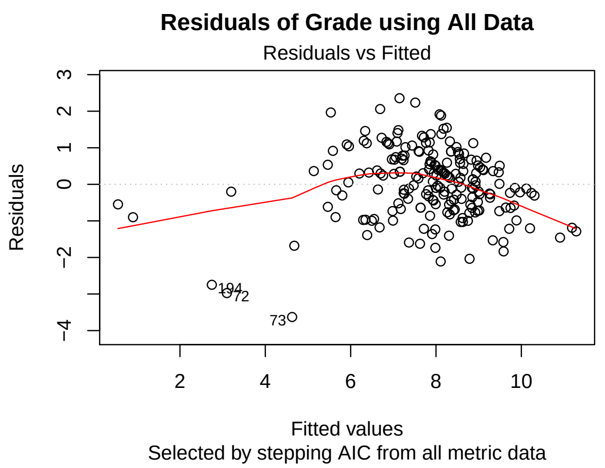
Is it possible to explain grades by means of metrics?

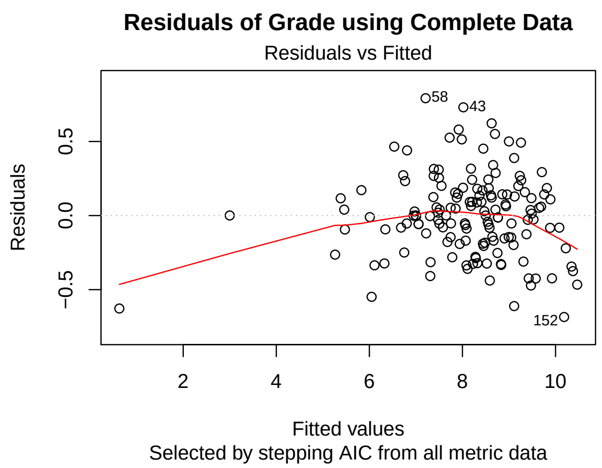
This will be done using linear regression and partial least squares regression. The data used for this chapter, unlike with ANOVA, has not had columns with risk factors (percentile and Alves) divided by their totals, because this reduces the standard deviation below 1 (since they become percentages), and partial least squares regression can not use data columns with a standard deviation below 1.

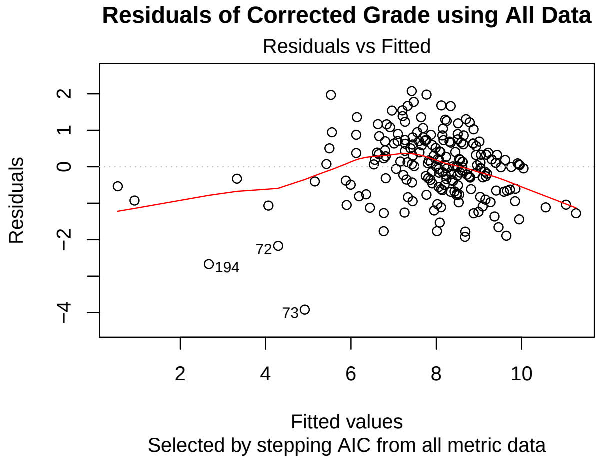
3.0.1 Performing Linear Regression

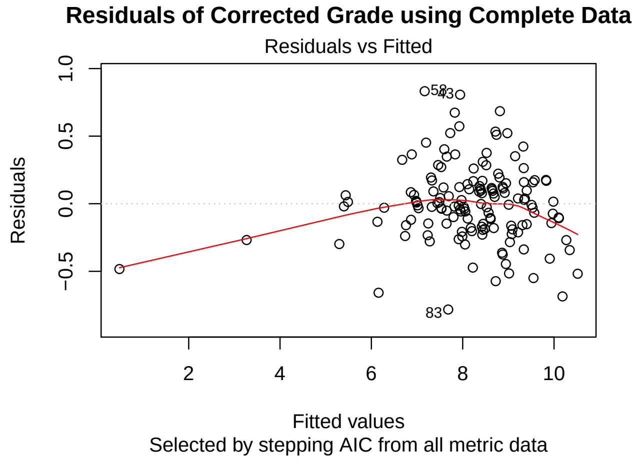
This linear regression features 8 models. These models consist of all 4 quality marks, using both all data and complete data. All of these models started as a linear combination of all metric-derived data columns, and were then stepped down (i.e. incrementally removing data columns) for a better model, as indicated by Akaike's An Information Criterion.

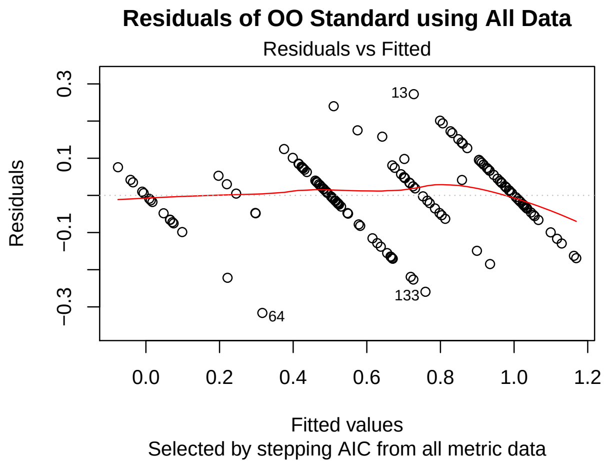
The figures below show a visual representation of the model, displayed as a red line, with the residuals (the difference of each data point from the predicted value, displayed by ordering the resultant variable ascendingly). This is an indicator of how well the chosen model matches the data it is trying to predict.

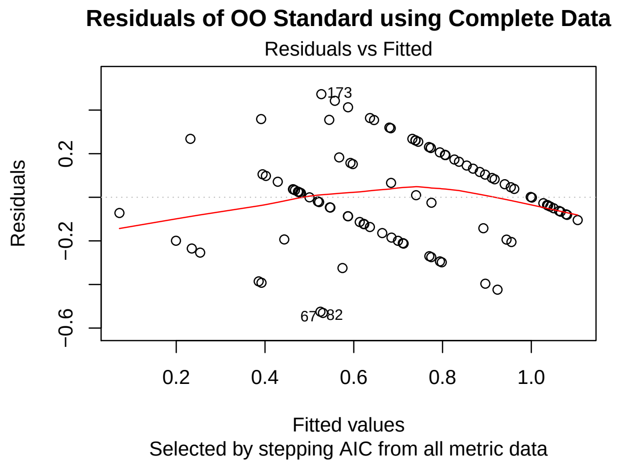


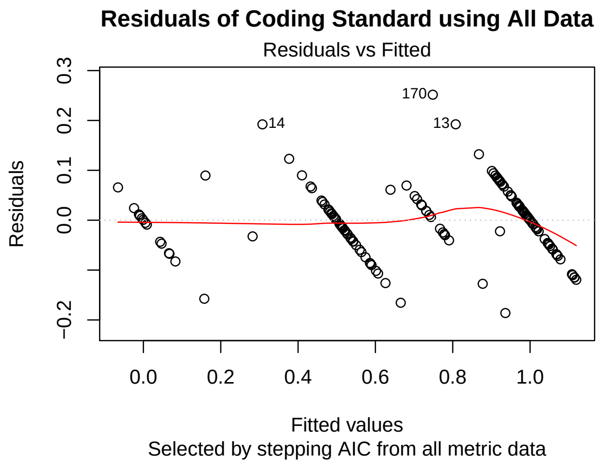


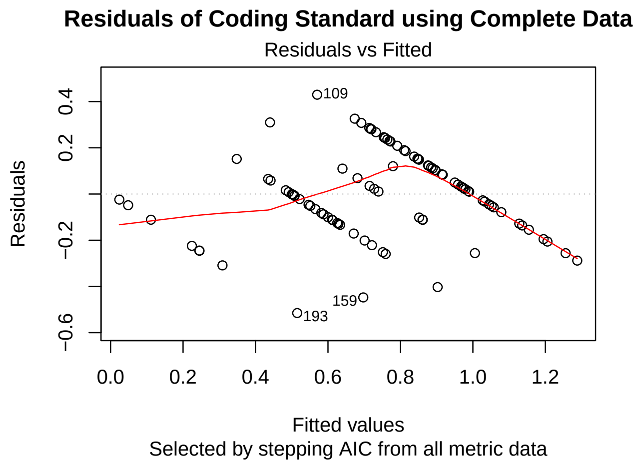












What can be seen in these tables, is that none of the models seem to conform particularly well to the data they are modelling. All grade models are clouds of points that seem to be arbitrarily ordered relative to the model line. What can be seen is that the centre of each cluster is mostly placed on each line, but besides that these graphs look as if there is little correlations between metrics and grades.

Concerning the standards, the results are also problematic, as these models show that the predicted value is not continuous. Instead, the residuals form diagonal lines at regular intervals, that correspond to values 0, 0.25, 0.5, 0.75 and 1. However, what is positive about that is that as long as the model is accurate within 0.125 of the real value, rounding to the next multiple of

0.25 is an accurate prediction of the grade for either standard. Unfortunately, the residuals are often higher, making this approach inviable, but a different approach might be able to predict these grades.

Either way, it is usually not sufficient to just look at the residuals graph to determine the goodness of fit. For that, the corrected R-squared is used, a percentage that indicates the amount of variance in the resultant variable is explained by the linear model. In the following table, the models from the graphs above are shown with the relevant R-squared.

Resultant Variable	Dataset	Explained Variance *Corrected R-Squared*
Grade	All Data	53.27 %
	Complete Data	83.98 %
Corrected Grade	All Data	55.14 %
	Complete Data	83.54 %
OO Standard	All Data for RPG, CardGame	66.49 %
	Complete Data for RPG, CardGame	34.53 %
Coding Standard	All Data for RPG, CardGame	79.68 %
	Complete Data for RPG, CardGame	50.49 %

The graphs above gave the idea that the models were not very accurate. The corrected R-squared forms a different conclusion. With nearly every model managing at least 50 % of variance explained, with a maximum of almost 84 %, there is a large indication that this data with this technique can use metrics to predict grades. While metric data does not explain everything, it is shown to explain more than 50 % in this scenario, which can no longer be attributed to just chance.

At least, that would be the case if this result was indisputable. That is not the case. The assumption for this chapter is that explaining grades is not possible, which is why it is necessary to explore the possibility that these models are overfitted, and that future data may not show the same level of variance explained. This is also just one study, that did not carefully control its conditions, which means that this should be treated as a single observation, not a study. These results only *suggest* that metrics explain grades.

Another shortcoming the poorer performance when using All Data. Failures, which are not included in Complete Data, seem to make the model worse, which is expected, as more successful groups means more homogeneous data and thus fewer outliers, but not desirable, because any application of this technology will be expected to produce all grades - not just passing ones.

3.0.2 Exploring Overfitting in Linear Models

The problem of overfitting relates to the number of data columns that predict a parameter. If each parameter has one data column, the result is a matrix equation that will always have a solution, provided all data columns are independent. This means that a model with one independent parameter per variable will always have 100% of variance explained.

The goal therefore is not to arrive at 100% variance explained, but to explain the largest amount of variance with the smallest amount of parameters, as the fewer parameters there are, the higher chance that its parameters are truly correlated with the resultant variable. Because of this, it is desirable to find models with fewer parameters than the models above.

Because of the nature of research, the requirement to use data with a standard deviation higher than 1 was discovered before the scaling was undone. This means that there is data available from models that were made then, which did not include these risk parameters. This provides an opportunity to evaluate the improvement achieved by using these risk parameters. Their corrected R-squared, and the amount of variance explained by the addition of those parameters are found in the table below:

Resultant Variable	Dataset	Explained Variance *Corrected R-Squared*	Risk Difference
Grade	All Data	34.94 %	18.33 %
	Complete Data	36.13 %	17.14 %
Corrected Grade	All Data	36.28 %	16.99 %
	Complete Data	34.85 %	18.42 %
OO Standard	All Data for RPG, CardGame	27.76 %	25.51 %
	Complete Data for RPG, CardGame	42.23 %	11.04 %
Coding Standard	All Data for RPG, CardGame	31.26 %	22.01 %
	Complete Data for RPG, CardGame	44.35 %	8.92 %

What can be seen here is that by adding this data the amount of variance explained went up. However, what is not displayed is the amount of parameters in each model. Because for validity a lower amount of parameters is more desirable, these models are compared on the amount of predictors. In the table below, the models, the number of data points they are based on, the number of data columns used in the model, the number of predictors per data point for models that used the Percentile Risk and Alves Risk data, and the same for the models that did not.

Dataset	Cases	Variables	Predictors (with Risks)	Predictors (without Risks)
All Data				
Grade	205	78	1 in 2.6	1 in 7.6
Corrected Grade	205	79	1 in 2.6	1 in 7.9
OO Standard	152	113	1 in 1.3	1 in 8.9
Coding Standard	152	121	1 in 1.3	1 in 6.3
Complete Cases				
Grade	147	110	1 in 1.3	1 in 4.6
Corrected Grade	147	107	1 in 1.4	1 in 4.5
OO Standard	98	28	1 in 3.5	1 in 2.8
Coding Standard	98	33	1 in 3	1 in 2.9

These models that use risks may explain a lot of variance, but they do so by using a lot more parameters, hence they are also more overfitted. The exceptions are OO Standard and Coding Standard for complete cases. To assess the rate of overfitting even better, these models are compared on how much variance they manage to explain relative to their degree of overfitting, by dividing the corrected R-squared by the number of predictors per case for a trade off between fit and overfitting:

Dataset	Cases	Variance / Predictor / Case (with Risks)	Variance / Predictor / Case (without Risks)
All Data			
Grade	205	1.4	2.65
Corrected Grade	205	1.43	2.86
OO Standard	152	0.89	2.48
Coding Standard	152	1	1.98
Complete Cases			
Grade	147	1.12	1.66
Corrected Grade	147	1.15	1.55
OO Standard	98	1.21	1.18
Coding Standard	98	1.5	1.28

In this comparison, the models with risks generally perform worse than the models that do not use risks. Based on the relative power of each predictor, they perform about half as well as the previous variants on all data. For complete cases it's about even. Because of this, it seems like a good idea to try and find models with around 30 parameters, but using the new data set to see how well those less overfitted models perform:

Dataset	Cases	Explained Variance	Predictors	Per Case	Variance / Predictor / Case
All Data					
Grade	205	42.45 %	18	1 in 11.4	4.83
Corrected Grade	205	44.54 %	23	1 in 8.9	3.97
OO Standard	152	43.04 %	15	1 in 10.1	4.36
Coding Standard	152	44.37 %	18	1 in 8.4	3.75
Complete Cases					
Grade	147	29.63 %	15	1 in 9.8	2.9
Corrected Grade	147	30.6 %	13	1 in 11.3	3.46
OO Standard	98	54.08 %	31	1 in 3.2	1.71
Coding Standard	98	66.58 %	50	1 in 2	1.31

These models were created by using the same stepping technique as the first models, this time by incrementally adding data columns until an optimal value for Akaike's An Information Criterion (AIC) is found. Besides the model of Coding standard with risks, all of these models extract more variance explained from fewer relative predictors. With most models having 10 predictors per variable, they are also not very overfitted.

That means that these results of around 42 % variance explained are probably a good estimate of what part of the grade can be explained. Since this value is not higher than 50 %, the indication is that grades can not be predicted from these metrics data using linear regression.

3.0.3 Model Parameters

One thing that can still be done with the calculated models is listing which model parameters are most informative. Although the setup of this research does not allow binding conclusions, giving suggestions for which metrics can be useful for follow up research can still improve potential future research. Because the models with few parameters seem to be the most informative, they will serve as the basis for this evaluation of relevant parameters. The most relevant parameters are:

For Grade:

- Alves Very High Risk Methods per Class
- Minimum Lack of Cohesion of Methods
- Median Fan-out.

For Corrected Grade:

- Total Number of Methods
- Alves Moderate Risk Fan-out

For OO Standard

- Percentile Very High Risk Fan-out
- Alves Moderate Risk Number of Methods
- Maximum Lack of Cohesion of Methods

For Coding Standard

- Total Number of Methods
- Mean Method Lines Of Code
- Maximum Lack of Cohesion of Methods
- Mean Number Of Fields
- Median Fan-out

These parameters were found by calculating the intersection of the model parameters for the linear models created by incrementally adding data columns until an optimum AIC was found.

Most of these parameters are considered less than ideal if they are high. It is therefore good to note that a higher maximum lack of cohesion of methods, meaning that functionality is not nicely divided over classes, increases both OO standard and coding standard. Another reason to doubt the results found here, although admittedly model parameters do not have a very clear meaning on their own.

3.0.4 Conclusion

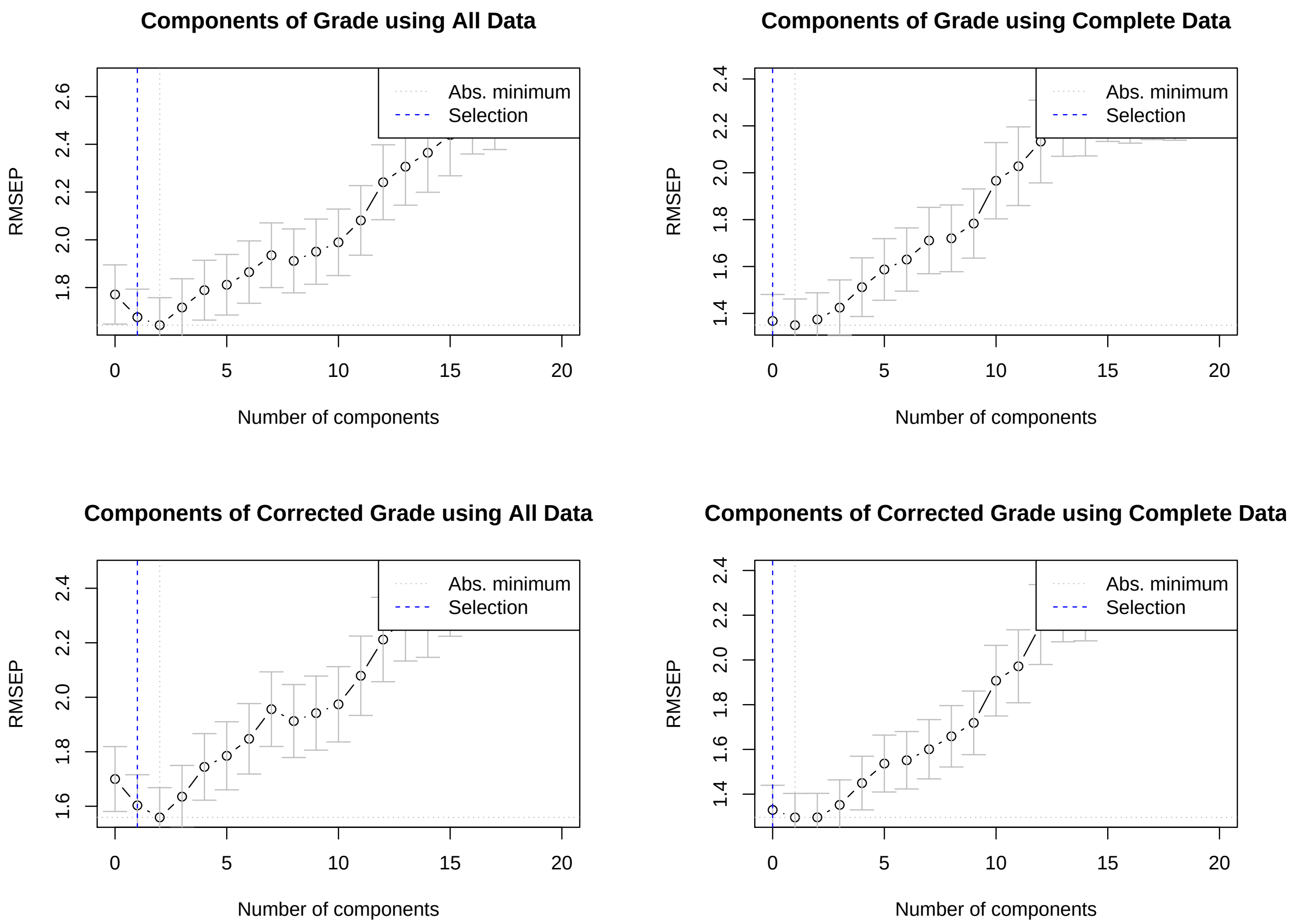
Initially this method of modelling seemed impressive, but closer inspection showed substantial overfitting in those models. A more restrained model managed to provide decent results, but not of the degree seen earlier. Because of this there seems to be no reason to assume that these linear models can indeed predict grades.

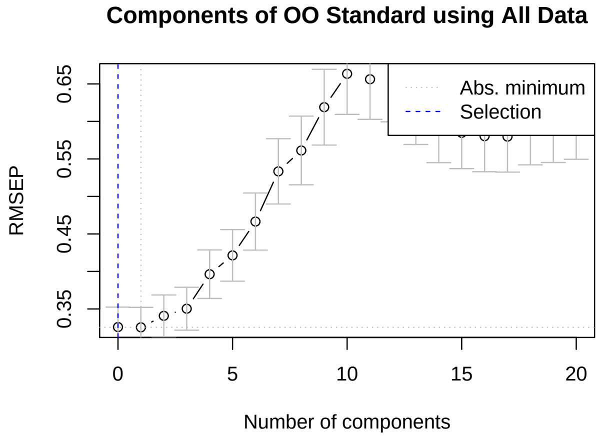
For future research, the recommendations are to properly design an experiment before attempting this problem again. A thorough theoretical approach should help overcome the difficulties in explaining the role of the various model parameters - a clear shortcoming of this study. It seems likely that at least adding different metrics than the ones used here, ideally more related to the programs under investigation will produce better results.

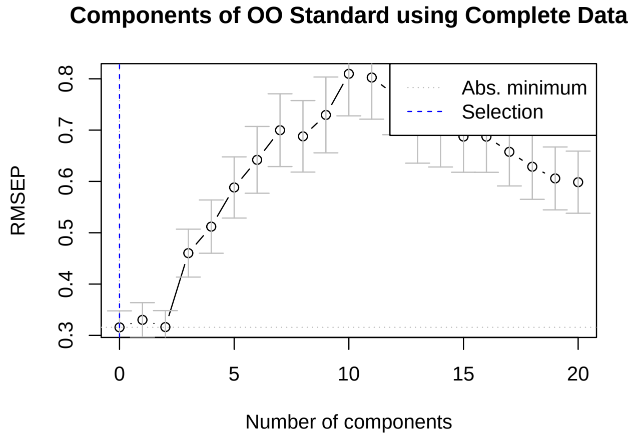
3.1 Partial Least Squares Regression

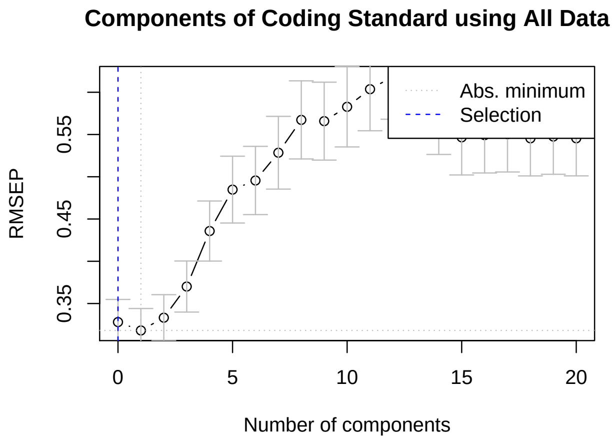
Partial Least Squares Regression is in a sense just like linear regression. However, unlike linear regression, it decomposes the model variables into components. These components are a linear combination of variables that form an orthogonal basis in the data. These components then receive a parameter similar to a linear model, and they constitute the eventual model.

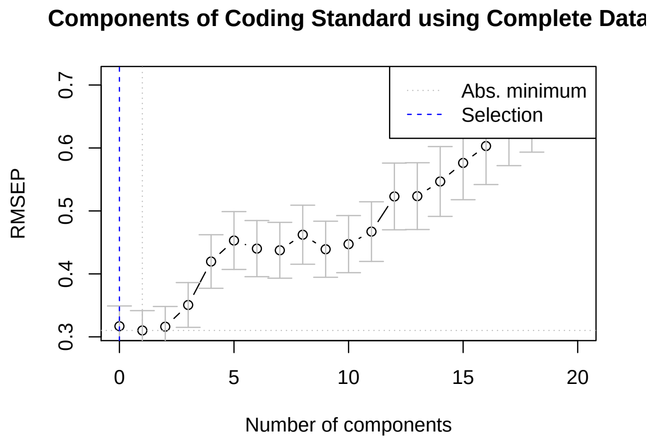
However, creating these components causes an issue: overdetermination. By recombining the model parameters they end up representing data variance so effectively that they quickly become overdetermined. To combat that, it is common to cross-validate each model, and use that information to select the number of components to use for a sensible model. The result can be seen here. (Mevik & Wehrens, 2019)











These scree plots plot the cross validation score for each successive component. A lower cross validation score is better. These are used to choose the number of components. Here, the one sigma approximation is used to select that number, which means taking the first component that is within one standard deviation of the cross validation score from the global optimum.

What can be seen is that these models barely improve their cross validation scores from the first component. That means that they rely very heavily on using this exact data, and do not generalize across different selections of data. This implies that any linear connection discovered here is not shared with the other data points. In other words: there is no linear relation between grades and metrics.

The only exception to this rule are 2 models, both based on all data, the largest data set, both allowed to use exactly 1 component, and both based on grade data, namely Grade and Corrected Grade. All other models stop at the intercept of the model - a constant. Here are their variance results:

Resultant Variable	Dataset	X-Variance	Y-Variance
Grade	All Data	25.17 %	14.47 %
Corrected Grade	All Data	24.73 %	15.00 %

Both these models manage to explain 15 % variance in the grade using 25 % of variance in the data. Any higher would not be valid following the results of the cross validation. As a result, this technique indicates that of the metric data only 25% may be used to explain grade results without loss of accuracy, and with that 25% of the data, only 15% of the grade can be explained.

As mentioned, the result that cross correlation scores do not increase show that the linear connection desired is not present. The fact that the results for grade using all data do produce one valid component, may well be caused by the fact that (Corrected) Grade using All Data has the most data points of all data sets. This result may well be the central limit theorem in effect. The fact that with complete data these same data sets results in zero components is consistent with that hypothesis.

3.2 Conclusion

In this section, an attempt was made to answer the following research question:

Is it possible to explain grades by means of metrics?

The hypothesis was that this should not be possible, because these metrics are not treated as goals of their own, but as values that may be used as indicators for desirable properties. As an example from the ANOVA results, the use of inheritance differed between projects, but this was because some assignments were designed to use inheritance, and some were not.

To test this hypothesis, using the functions of R, models were generated using all data points, and with one of those models 80% of the grade was explained. However, when compared to other models, that model was highly overfitted, meaning that the validity of correlation is pulled into question. However, the cost of fixing overfitting was lower explained variance.

Although the models produced a good degree of fit for each variable and data set, this was not entirely convincing when looking at the graphs. The grade data was a cloud of points in the range of passing grades; the standards a few lines around the few levels those value achieved. Where the standard grades gave the impression that another technique might be able to predict them, the grade data did not.

To try and address at least the overfitting, some new models were generated. While these methods did reach a parameter to result ratio of about 1 to 10, much better than the 1 to 2 it previously had, their rate of explanation dropped below 50 %, meaning that linear regression did not produce a model that was accurate without being overfitted. As a result, linear regression was deemed unable to refute the hypothesis and show that metrics predict grades.

After the inability of linear regression to refute the hypothesis, another attempt was made by using partial least squares regression. This technique attempts to calculate the most accurate model possible using linear combinations of data called components as factors. However, only two models produced components that improved the cross validation scores that determine how many components a model should use.

This result means that the data is very sensitive to removing points. In other words: The models made here appear not to have deep linear connections between metrics and grades. This also suggests that repeating this research will likely result in a similar lack of correlation between grades and metrics, which is supportive of the original hypothesis.

Nevertheless, when using all data, the largest data set, and either the grade or the corrected grade (values that differ very little), using 25 % of the variance in the metrics, 15 % of the variance in the grade is explained. This is much less than the linear models, which reach about 45%. Either way, the hypothesis that grades do not explain metrics can not be rejected.

Because of this, the answer to the question:

Is it possible to explain grades by means of metrics?

remains: It should not be possible to explain grades by means of metrics. The performed analyses do not indicate a real linear relation between them.

Chapter 4

Conclusion

This thesis tried to answer the following research questions, and did so by testing their corresponding hypotheses. Below those hypotheses, the found answers are included:

Is there an effect of groups on project grades?

Hypothesis There should be an effect of groups on project grades.

Result There is an effect of groups on project grades - groups received different grades.

Is there an effect of groups on project metrics?

Hypothesis There should be an effect of groups on project metrics.

Result There is an effect of groups on project metrics - groups showed differences that indicated that they varied their program designs.

Is there an effect of teaching assistants on project grades?

Hypothesis There should be an effect of teaching assistants on project grades.

Result There is an effect of teaching assistants on project grades - although teaching assistants graded mostly consistently, they gave different grades for OO Standard and Coding Standard.

Is there an effect of teaching assistants on project metrics?

Hypothesis There should not be any effect of teaching assistants on project metrics.

Result No, there is no real effect of teaching assistants on project metrics, all effects seem sufficiently explained by attributing them to systemic error.

Is there an effect of assignments on project grades?

Hypothesis There should not be an effect of assignments on project grades.

Result There technically was an effect, as coding standard was different between assignments. To be precise, it increased between RPG and Card Game. Other grades were not significantly different however.

Is there an effect of assignments on project metrics?

Hypothesis There should be an effect of assignments on project metrics.

Result A large amount of metrics differed between projects, showing the differences between these assignments.

Is it possible to explain grades by means of metrics?

Hypothesis It should not be possible to explain grades by means of metrics.

Result It should not be possible to explain grades by means of metrics. The performed analyses do not indicate a real linear relation between them.

What these results show is that the design of the course is mostly consistent with what the course should be. Certain results such as a consistently lower coding standard grade when comparing two groups of teaching assistants could indicate the potential need to reassess these subjective grades in the grading system, and certain similarities between Card Game and Graph Editor might be cause to evaluate those assignments.

On the other hand, these results do not produce many clear contradictions with assumptions. In part that is because of the volume of results. The used metrics were not curated for their predictability, and with roughly 200 different metric aggregations, 200 distinct hypotheses needed to be formulated. Given the lack of experimental correlation with code properties of these metrics, and the obfuscation caused by the aggregation, this was not feasible.

In part, however, the contradictions were assumed to be caused by systematic issues, notably the differences of metrics between teaching assistants. In that case, arguments could be found in favour of the explanation that allowed attributing them to systematic error. In the case of OO standard, that contradiction still needs to be looked into.

Nevertheless, the overall image is that the results are consistent with the presented hypotheses. This means that the intuitions that lead to them are now confirmed once. However, it might be valuable to repeat this research for a later edition of the course, or for another course to see if they also hold in multiple courses.

4.1 Future Research

Given that during literature research no similar studies were found, it is likely sensible to repeat this research at least once, in which case most of this thesis can be reused. If this research is ever repeated, it may be interesting to enhance it with any of the following suggestions.

- Focus on the meaning of a specific metric and see what its influence is. The theoretical role of metrics is by design somewhat loosely connected to code, but if these metrics can be employed for directed analyses of specific properties that might improve the specificity of the above results.
- Focus on the effect of assignments on certain grades. Ideally this should be done in a course that registers quality for every assignment (which will in general be beneficial when repeating this research). The current data shows an effect, but can not determine if it is caused by students (or teaching assistants) getting used to the course environment, or by them actually improving.
- Attempt another technique to analyse the grades. One of the more promising results was the linear model on OO standard, which if combined with some manner of threshold setting that ensures results are one of 0, .25, .5, .75, or 1, could be somewhat accurate. If the same is applied, if applicable, to individual grade elements, rather than the entire grade, it may be possible to predict the grade.
- A shortcoming of this thesis is that it has difficulty motivating why the correlation between metrics and grades should exist, and that is most likely because the grade is not

sufficiently subdivided into relevant items. It may be helpful to try and create tests for these individual items to achieve a fully automated grading system. Exploring the grade models more closely might be a good idea.

- Those intending to explore the grade models will probably run into the need to test functionality. When that happens, it might be possible to, via one or more example programs, train a neural network to attempt certain actions, and see if that neural network is able to test that functionality for the programs in the course.

Bibliography

Abreu, F. B. & Melo, W. (1996). Evaluating the impact of object-oriented design on software quality. In *In proceedings of the 3 rd international software metrics symposium* (pp. 90–99).

Alves, T. L., Ypma, C., & Visser, J. (2010). Deriving metric thresholds from benchmark data. In *2010 ieee international conference on software maintenance* (pp. 1–10).

Bansal, M. & Agrawal, C. P. (2014). Critical analysis of object oriented metrics in software development. In *2014 fourth international conference on advanced computing communication technologies* (pp. 197–201).

Basili, V. R., Briand, L. C., & Melo, W. L. (1996). A validation of object-oriented design metrics as quality indicators. *IEEE Transactions on Software Engineering*, *22*(10), 751–761.

Boucher, A. & Badri, M. (2017). Predicting fault-prone classes in object-oriented software: an adaptation of an unsupervised hybrid som algorithm. In *2017 ieee international conference on software quality, reliability and security (qrs)* (pp. 306–317).

Checkstyle. (2020). Checkstyle - checkstyle 8.32. http://checkstyle.sourceforge.net/. Accessed: 2020-05-22.

Chidamber, S. R. & Kemerer, C. F. (1994). A metrics suite for object oriented design. *IEEE Transactions on Software Engineering*, *20*(6), 476–493.

Designite. (2019). Designite - reduce technical debt of your software. http://www.designite-tools.com/. Accessed: 2019-04-30.

Facebook, Inc. (2019). Infer static analyzer — infer — infer. http://fbinfer.com/. Accessed: 2019-04-30.

Mevik, B.-H. & Wehrens, R. (2019). Introduction to the pls package. Documentation for the R package pls. Relevant to R version 3.2.3.

PMD. (2019). Pmd. https://pmd.github.io/. Accessed: 2019-04-30.

Rod Hilton. (2019). Jasome (java source metrics) - object oriented metrics analyzer for java code. https://github.com/rodhilton/jasome. Accessed: 2019-04-30.

SpotBugs. (2019). Spotbugs. https://spotbugs.github.io/. Accessed: 2019-04-30.

Verzani, J. (2014a). Using r for introductory statistics. In *Using r for introductory statistics* (p. 404). Boca Raton, FL: CRC Press.

Verzani, J. (2014b). Using r for introductory statistics. In *Using r for introductory statistics* (p. 390). Boca Raton, FL: CRC Press.

Part I

Appendix 1 : Tables

4.2 ANOVA Result Tables

In the following tables, the results of the ANOVA analyses will be shown. In these tables, the p-value will be shown of performing ANOVA on a model that attempts to explain variance in the variable under **Variable** using the category listed above it.

Variable	**Group P-value**	**Project P-value**	**TA's (A) P-value**	**TA's (B) P-value**	**TA's (C) P-value**
Data Group					
Data	p-value	p-value	p-value	p-value	p-value

These tables tabulate the results of performing ANOVA using data column Data, which is part of Data Group, with the category it crosses with, reporting the p-value of that ANOVA. For example, in the Data Group 'Resultant Variable', there is the data column 'Corrected Grade'. In these tables, each p-value can be marked with one of the following coloured labels:

Significance level / Assumption	[1,0.1)	[0.1, 0.05)	[0.05, 0.01)	[0.01, 0.001)	[0.001, 0)
All Assumptions Met					
Homogeneous, not Normal					
No Assumptions Met					

These labels are intended to distinguish between different levels of significance for ANOVAs, as well as differentiating between which assumptions an ANOVA meets according to the Shapiro-Wilk normality test (indicated as 'Normal' above) and Levene's test for homogeneity of variance ('Homogeneous') To this end a colour is added to each cell to indicate the assumptions met, the opacity of the colour relates to the level of significance.

The astute reader will notice that there is no colour option for Normal, not Homogeneous. To arrive at this approach, some tests with the data were performed, and the results of those tests showed that such a situation does not occur in the data used in this thesis.

There will be two tables, each will list which data set its analyses are based on.

Table 4.1: ANOVA for 4 categories on all data

Variable	**Group P-value**	**Project P-value**	**TA's (A) P-value**	**TA's (B) P-value**	**TA's (C) P-value**
Resultant Variable					
OO Standard	0.002	0.218	< 0.001	< 0.001	< 0.001
Coding Standard	0.578	< 0.001	< 0.001	< 0.001	< 0.001
Grade	< 0.001	0.346	0.372	0.341	0.643
Corrected Grade	< 0.001	0.23	0.337	0.273	0.595
Total					
Project Number of Methods	0.382	< 0.001	0.122	0.108	0.004
Project Number of Types	0.287	< 0.001	0.041	0.032	0.001
Method Lines of Code (MLOC)					
Minimum	1	< 0.001	0.13	0.713	< 0.001
Maximum	0.957	< 0.001	0.22	0.506	0.458
Median	0.871	< 0.001	0.197	0.321	0.078
Mean	0.796	< 0.001	0.196	0.125	0.319
Variance	0.949	< 0.001	0.377	0.451	0.671
Standard Deviation	0.987	< 0.001	0.285	0.403	0.524
Percentile Low Risk	0.388	0.023	0.079	0.043	0.052
Percentile Medium Risk	0.917	< 0.001	0.374	0.208	0.087
Percentile High Risk	0.454	0.116	0.252	0.218	0.402
Percentile Very High Risk	0.002	0.635	0.85	0.778	0.944
Alves Low Risk	0.743	< 0.001	0.63	0.629	0.843
Alves Medium Risk	0.436	< 0.001	0.969	0.627	0.711
Alves High Risk	0.616	< 0.001	0.288	0.565	0.245
Alves Very High Risk	0.172	0.007	0.235	0.189	0.359
Cyclomatic Complexity (CC)					
Minimum	-	-	-	-	-
Maximum	0.986	< 0.001	0.009	0.116	0.028
Median	-	-	-	-	-
Mean	0.377	< 0.001	0.315	0.394	0.278
Variance	0.962	< 0.001	0.091	0.228	0.274
Standard Deviation	0.964	< 0.001	0.031	0.2	0.084
Percentile Low Risk	-	-	-	-	-
Percentile Medium Risk	0.328	0.306	0.816	0.272	0.517
Percentile High Risk	0.899	< 0.001	0.613	0.058	0.027
Percentile Very High Risk	0.015	0.008	0.747	0.797	0.786
Alves Low Risk	0.803	< 0.001	0.456	0.519	0.531
Alves Medium Risk	0.831	< 0.001	0.552	0.664	0.336
Alves High Risk	0.925	< 0.001	0.886	0.919	0.975
Alves Very High Risk	0.01	< 0.001	0.059	0.144	0.128

Variable	Group P-value	Project P-value	TA's (A) P-value	TA's (B) P-value	TA's (C) P-value
Parameter Count (PC)					
Minimum	0.903	0.453	0.676	0.676	0.926
Maximum	0.619	< 0.001	0.501	0.887	0.058
Median	1	< 0.001	0.355	0.547	0.196
Mean	0.997	< 0.001	0.563	0.737	0.567
Variance	0.339	< 0.001	0.323	0.333	0.503
Standard Deviation	0.366	< 0.001	0.597	0.714	0.698
Percentile Low Risk	1	< 0.001	0.709	0.852	0.6
Percentile Medium Risk	-	-	-	-	-
Percentile High Risk	1	< 0.001	0.879	0.917	0.498
Percentile Very High Risk	1	< 0.001	0.758	0.703	0.514
Alves Low Risk	0.281	0.003	0.382	0.331	0.647
Alves Medium Risk	0.148	0.975	0.427	0.992	0.204
Alves High Risk	0.489	0.069	0.718	0.21	0.343
Alves Very High Risk	0.477	< 0.001	0.24	0.433	0.207
Number Of Fields (NOF)					
Minimum	0.846	0.816	0.397	0.22	0.542
Maximum	0.432	0.154	0.538	0.643	0.742
Median	0.026	0.044	0.238	0.283	0.515
Mean	0.005	0.131	0.523	0.477	0.647
Variance	0.596	0.677	0.523	0.505	0.809
Standard Deviation	0.039	0.493	0.573	0.641	0.708
Percentile Low Risk	0.056	< 0.001	0.603	0.453	0.442
Percentile Medium Risk	0.957	< 0.001	0.25	0.706	< 0.001
Percentile High Risk	0.3	0.217	0.073	0.058	0.014
Percentile Very High Risk	< 0.001	0.279	0.139	0.507	0.077
Alves Low Risk	0.006	0.003	0.214	0.543	0.099
Alves Medium Risk	0.461	0.209	0.028	0.058	0.052
Alves High Risk	0.219	0.1	0.105	0.453	0.193
Alves Very High Risk	0.04	0.031	0.858	0.554	0.603
Number Of Private Fields (NOPF)					
Minimum	0.903	0.453	0.277	0.063	0.238
Maximum	< 0.001	0.793	0.093	0.141	0.161
Median	0.883	0.365	0.072	0.015	0.074
Mean	0.115	0.372	0.034	0.006	0.032
Variance	0.001	0.595	0.13	0.031	0.123
Standard Deviation	< 0.001	0.586	0.057	0.03	0.067
Percentile Low Risk	-	-	-	-	-
Percentile Medium Risk	-	-	-	-	-
Percentile High Risk	-	-	-	-	-
Percentile Very High Risk	-	-	-	-	-
Alves Low Risk	0.027	0.146	0.307	0.053	0.163
Alves Medium Risk	0.728	0.203	0.234	0.255	0.26
Alves High Risk	0.689	0.144	0.83	0.776	0.866
Alves Very High Risk	0.005	0.022	0.022	0.004	0.016

Variable	Group P-value	Project P-value	TA's (A) P-value	TA's (B) P-value	TA's (C) P-value
Number Of Methods (NOM)					
Minimum	0.575	0.45	0.857	0.914	0.981
Maximum	0.471	< 0.001	0.635	0.576	0.106
Median	0.041	0.048	0.663	0.652	0.419
Mean	0.002	< 0.001	0.751	0.717	0.805
Variance	0.306	< 0.001	0.641	0.629	0.642
Standard Deviation	0.176	< 0.001	0.727	0.679	0.468
Percentile Low Risk	0.002	0.045	0.869	0.723	0.795
Percentile Medium Risk	0.24	0.006	0.988	0.765	0.908
Percentile High Risk	0.04	0.59	0.225	0.33	0.096
Percentile Very High Risk	< 0.001	0.458	0.167	0.254	0.193
Alves Low Risk	0.013	0.002	0.103	0.129	0.082
Alves Medium Risk	0.196	0.255	0.192	0.177	0.288
Alves High Risk	0.107	0.081	0.649	0.546	0.675
Alves Very High Risk	0.009	0.046	0.274	0.518	0.457
Number Of Private Methods (NOPM)					
Minimum	0.856	< 0.001	0.072	0.282	0.053
Maximum	0.582	< 0.001	0.525	0.286	0.08
Median	0.016	0.105	0.344	0.41	0.148
Mean	< 0.001	< 0.001	0.849	0.814	0.703
Variance	0.327	< 0.001	0.718	0.429	0.502
Standard Deviation	0.11	< 0.001	0.841	0.401	0.34
Percentile Low Risk	0.006	0.897	0.643	0.492	0.402
Percentile Medium Risk	0.536	0.28	0.718	0.856	0.645
Percentile High Risk	0.208	0.137	0.294	0.175	< 0.001
Percentile Very High Risk	< 0.001	0.486	0.207	0.362	0.296
Alves Low Risk	0.045	< 0.001	0.006	0.025	0.002
Alves Medium Risk	0.375	0.166	0.156	0.167	0.216
Alves High Risk	0.083	0.002	0.332	0.405	0.325
Alves Very High Risk	< 0.001	< 0.001	0.083	0.534	0.15
Type Lines of Code (TLOC)					
Minimum	0.921	< 0.001	0.133	0.545	< 0.001
Maximum	0.041	0.194	0.246	0.393	0.12
Median	< 0.001	0.115	0.239	0.208	0.431
Mean	< 0.001	0.438	0.224	0.229	0.321
Variance	0.004	0.069	0.019	0.01	0.011
Standard Deviation	0.015	0.117	0.047	0.035	0.022
Percentile Low Risk	< 0.001	0.684	0.566	0.558	0.82
Percentile Medium Risk	0.011	0.7	0.137	0.36	0.064
Percentile High Risk	< 0.001	0.324	0.267	0.078	0.058
Percentile Very High Risk	< 0.001	0.996	0.186	0.259	0.389
Alves Low Risk	0.166	0.082	0.543	0.676	0.489
Alves Medium Risk	0.245	0.137	0.551	0.507	0.667
Alves High Risk	0.368	0.162	0.628	0.438	0.516
Alves Very High Risk	0.032	0.611	0.083	0.023	0.039

Variable	Group P-value	Project P-value	TA's (A) P-value	TA's (B) P-value	TA's (C) P-value
Weighted Methods per Class (WMC)					
Minimum	0.322	0.126	0.706	0.669	0.561
Maximum	0.082	< 0.001	0.27	0.837	0.159
Median	0.039	0.002	0.987	0.993	0.999
Mean	< 0.001	0.702	0.616	0.874	0.733
Variance	0.002	0.315	0.024	0.219	0.074
Standard Deviation	0.004	0.215	0.076	0.437	0.111
Percentile Low Risk	< 0.001	0.71	0.891	0.839	0.91
Percentile Medium Risk	0.068	0.116	0.588	0.373	0.542
Percentile High Risk	0.198	0.604	0.996	0.982	0.933
Percentile Very High Risk	< 0.001	0.889	0.196	0.124	0.265
Alves Low Risk	0.072	0.227	0.717	0.716	0.649
Alves Medium Risk	0.224	0.47	0.613	0.485	0.771
Alves High Risk	0.387	0.517	0.713	0.785	0.652
Alves Very High Risk	0.013	0.527	0.175	0.132	0.146
Number of Children (NC)					
Minimum	-	-	-	-	-
Maximum	0.755	0.004	0.019	0.023	0.037
Median	0.018	0.238	0.115	0.196	0.253
Mean	0.977	< 0.001	0.41	0.506	0.493
Variance	0.976	< 0.001	0.12	0.122	0.329
Standard Deviation	0.875	< 0.001	0.05	0.067	0.075
Percentile Low Risk	-	-	-	-	-
Percentile Medium Risk	-	-	-	-	-
Percentile High Risk	1	< 0.001	0.569	0.835	0.139
Percentile Very High Risk	1	< 0.001	0.569	0.835	0.139
Alves Low Risk	0.547	< 0.001	0.323	0.237	0.539
Alves Medium Risk	0.418	< 0.001	0.194	0.127	0.16
Alves High Risk	0.856	0.768	0.247	0.293	0.254
Alves Very High Risk	0.853	< 0.001	0.785	0.815	0.947
Depth of Inheritance Tree (DIT)					
Minimum	-	-	-	-	-
Maximum	0.581	< 0.001	0.299	0.36	0.132
Median	0.969	< 0.001	0.632	0.717	0.831
Mean	0.978	< 0.001	0.517	0.593	0.6
Variance	0.927	< 0.001	0.656	0.775	0.715
Standard Deviation	0.747	< 0.001	0.271	0.396	0.177
Percentile Low Risk	1	< 0.001	0.215	0.419	0.053
Percentile Medium Risk	-	-	-	-	-
Percentile High Risk	0.914	< 0.001	0.498	0.52	0.676
Percentile Very High Risk	1	< 0.001	0.494	0.773	0.125
Alves Low Risk	0.064	0.042	0.139	0.118	0.214
Alves Medium Risk	0.002	0.266	0.164	0.101	0.132
Alves High Risk	0.071	0.494	0.508	0.57	0.646
Alves Very High Risk	0.631	0.022	0.512	0.477	0.602

Variable	Group P-value	Project P-value	TA's (A) P-value	TA's (B) P-value	TA's (C) P-value
Lack of Cohesion Of Methods (LCOM)					
Minimum	0.595	0.956	0.338	0.215	0.534
Maximum	0.999	< 0.001	0.396	0.65	0.298
Median	0.106	0.405	0.066	0.074	0.221
Mean	0.344	< 0.001	0.53	0.605	0.799
Variance	0.937	< 0.001	0.466	0.547	0.441
Standard Deviation	0.948	< 0.001	0.423	0.583	0.392
Percentile Low Risk	0.992	< 0.001	0.706	0.823	0.275
Percentile Medium Risk	0.995	< 0.001	0.016	0.11	0.017
Percentile High Risk	1	< 0.001	0.476	0.961	0.13
Percentile Very High Risk	0.016	0.132	0.363	0.359	0.632
Alves Low Risk	0.181	< 0.001	0.047	0.156	0.123
Alves Medium Risk	0.049	0.156	0.095	0.192	0.086
Alves High Risk	0.878	< 0.001	0.067	0.233	0.077
Alves Very High Risk	0.519	< 0.001	0.509	0.453	0.662
Fan-in (FANIN)					
Minimum	0.345	0.541	0.193	0.173	< 0.001
Maximum	0.435	< 0.001	0.152	0.113	0.01
Median	0.015	< 0.001	0.146	0.188	0.072
Mean	0.119	< 0.001	0.96	0.849	0.913
Variance	0.768	< 0.001	0.236	0.095	0.003
Standard Deviation	0.234	< 0.001	0.169	0.097	0.005
Percentile Low Risk	0.136	< 0.001	0.217	0.262	0.1
Percentile Medium Risk	1	< 0.001	0.855	0.715	0.902
Percentile High Risk	0.741	< 0.001	0.14	0.098	0.004
Percentile Very High Risk	0.422	< 0.001	0.938	0.7	0.493
Alves Low Risk	< 0.001	< 0.001	0.511	0.211	0.062
Alves Medium Risk	< 0.001	0.106	0.64	0.506	0.453
Alves High Risk	0.013	< 0.001	0.808	0.247	0.308
Alves Very High Risk	0.693	< 0.001	0.01	0.239	0.033
Fan-out (FANOUT)					
Minimum	0.885	0.013	0.764	0.908	0.262
Maximum	0.408	< 0.001	0.614	0.834	0.589
Median	0.13	< 0.001	0.357	0.456	0.65
Mean	0.119	< 0.001	0.96	0.849	0.913
Variance	0.119	< 0.001	0.378	0.87	0.464
Standard Deviation	0.163	< 0.001	0.199	0.725	0.179
Percentile Low Risk	0.969	< 0.001	0.938	0.556	0.584
Percentile Medium Risk	1	< 0.001	0.976	0.641	0.566
Percentile High Risk	0.935	< 0.001	0.816	0.946	0.917
Percentile Very High Risk	0.104	0.003	0.725	0.433	0.348
Alves Low Risk	0.234	< 0.001	0.759	0.677	0.067
Alves Medium Risk	0.867	< 0.001	0.936	0.99	0.17
Alves High Risk	0.335	0.002	0.135	0.742	0.005
Alves Very High Risk	0.048	0.215	0.737	0.757	0.8

Table 4.2: ANOVA for 4 categories on complete data

Variable	**Group P-value**	**Project P-value**	**TA's (A) P-value**	**TA's (B) P-value**	**TA's (C) P-value**
Resultant Variable					
OO Standard	< 0.001	0.609	0.021	0.021	0.021
Coding Standard	0.892	< 0.001	0.008	0.008	0.008
Grade	< 0.001	0.235	0.995	0.846	0.967
Corrected Grade	< 0.001	0.069	0.979	0.758	0.943
Total					
Project Number of Methods	0.636	< 0.001	0.186	0.006	0.004
Project Number of Types	0.567	< 0.001	0.038	< 0.001	< 0.001
Method Lines of Code (MLOC)					
Minimum	1	< 0.001	0.075	0.658	< 0.001
Maximum	0.942	< 0.001	0.659	0.815	0.851
Median	0.828	< 0.001	0.154	0.172	0.105
Mean	0.867	< 0.001	0.054	0.01	0.041
Variance	0.8	< 0.001	0.672	0.622	0.887
Standard Deviation	0.973	< 0.001	0.569	0.404	0.686
Percentile Low Risk	0.736	0.024	0.08	0.036	0.027
Percentile Medium Risk	0.996	< 0.001	0.18	0.03	0.048
Percentile High Risk	0.188	0.117	0.601	0.567	0.678
Percentile Very High Risk	0.005	0.807	0.546	0.528	0.725
Alves Low Risk	0.555	< 0.001	0.373	0.499	0.674
Alves Medium Risk	0.251	< 0.001	0.421	0.661	0.506
Alves High Risk	0.484	< 0.001	0.218	0.358	0.189
Alves Very High Risk	0.048	0.075	0.325	0.263	0.44
Cyclomatic Complexity (CC)					
Minimum	-	-	-	-	-
Maximum	0.972	< 0.001	0.379	0.767	0.464
Median	-	-	-	-	-
Mean	0.467	< 0.001	0.275	0.142	0.194
Variance	0.786	< 0.001	0.488	0.544	0.734
Standard Deviation	0.901	< 0.001	0.403	0.424	0.428
Percentile Low Risk	-	-	-	-	-
Percentile Medium Risk	0.536	0.335	0.896	0.524	0.639
Percentile High Risk	0.959	0.022	0.697	0.462	0.091
Percentile Very High Risk	0.04	0.049	0.219	0.609	0.46
Alves Low Risk	0.707	< 0.001	0.477	0.234	0.298
Alves Medium Risk	0.676	< 0.001	0.591	0.316	0.147
Alves High Risk	0.582	< 0.001	0.925	0.817	0.965
Alves Very High Risk	0.002	0.007	0.252	0.364	0.432

Variable	**Group P-value**	**Project P-value**	**TA's (A) P-value**	**TA's (B) P-value**	**TA's (C) P-value**
Parameter Count (PC)					
Minimum	0.489	0.37	0.782	0.782	0.965
Maximum	0.639	< 0.001	0.383	0.602	0.084
Median	1	< 0.001	0.023	0.433	0.07
Mean	0.988	< 0.001	0.1	0.336	0.16
Variance	0.521	< 0.001	0.324	0.273	0.36
Standard Deviation	0.515	< 0.001	0.466	0.575	0.515
Percentile Low Risk	1	< 0.001	0.123	0.552	0.228
Percentile Medium Risk	-	-	-	-	-
Percentile High Risk	1	< 0.001	0.119	0.864	0.185
Percentile Very High Risk	1	< 0.001	0.292	0.976	0.406
Alves Low Risk	0.302	0.002	0.828	0.775	0.957
Alves Medium Risk	0.536	0.448	0.958	0.343	0.144
Alves High Risk	0.564	0.044	0.53	0.396	0.584
Alves Very High Risk	0.419	< 0.001	0.982	0.428	0.561
Number Of Fields (NOF)					
Minimum	0.522	0.608	0.438	0.052	0.211
Maximum	0.311	0.416	0.689	0.837	0.869
Median	0.181	0.033	0.089	0.154	0.231
Mean	0.026	0.14	0.557	0.349	0.514
Variance	0.326	0.576	0.768	0.697	0.923
Standard Deviation	0.076	0.785	0.789	0.817	0.843
Percentile Low Risk	0.094	< 0.001	0.114	0.72	0.312
Percentile Medium Risk	0.962	< 0.001	0.029	0.948	< 0.001
Percentile High Risk	0.166	0.539	0.533	0.209	0.058
Percentile Very High Risk	0.044	0.176	0.847	0.484	0.41
Alves Low Risk	0.077	0.003	0.71	0.139	0.199
Alves Medium Risk	0.527	0.078	0.289	0.297	0.397
Alves High Risk	0.155	0.249	0.162	0.332	0.196
Alves Very High Risk	0.079	0.063	0.835	0.103	0.315
Number Of Private Fields (NOPF)					
Minimum	0.489	0.37	0.144	< 0.001	0.008
Maximum	< 0.001	0.641	0.026	0.028	0.023
Median	0.517	0.286	0.04	< 0.001	0.002
Mean	0.023	0.264	0.024	< 0.001	0.001
Variance	< 0.001	0.448	0.094	0.001	0.009
Standard Deviation	< 0.001	0.438	0.02	0.002	0.004
Percentile Low Risk	-	-	-	-	-
Percentile Medium Risk	-	-	-	-	-
Percentile High Risk	-	-	-	-	-
Percentile Very High Risk	-	-	-	-	-
Alves Low Risk	0.042	0.416	0.233	0.006	0.033
Alves Medium Risk	0.749	0.524	0.027	0.104	0.094
Alves High Risk	0.295	0.028	0.746	0.891	0.834
Alves Very High Risk	0.015	0.058	0.002	< 0.001	< 0.001

Variable	Group P-value	Project P-value	TA's (A) P-value	TA's (B) P-value	TA's (C) P-value
Number Of Methods (NOM)					
Minimum	0.346	0.309	0.576	0.505	0.832
Maximum	0.436	< 0.001	0.689	0.263	0.167
Median	0.246	0.047	0.167	0.892	0.19
Mean	0.041	< 0.001	0.79	0.838	0.775
Variance	0.146	< 0.001	0.424	0.357	0.496
Standard Deviation	0.182	< 0.001	0.37	0.251	0.224
Percentile Low Risk	0.073	0.112	0.981	0.865	0.856
Percentile Medium Risk	0.192	0.051	0.688	0.22	0.443
Percentile High Risk	0.013	0.94	0.035	0.06	0.017
Percentile Very High Risk	< 0.001	0.757	0.057	0.092	0.075
Alves Low Risk	0.109	< 0.001	0.62	0.543	0.442
Alves Medium Risk	0.344	0.034	0.476	0.09	0.185
Alves High Risk	0.548	0.038	0.263	0.19	0.331
Alves Very High Risk	0.032	0.158	0.004	0.071	0.026
Number Of Private Methods (NOPM)					
Minimum	0.865	< 0.001	0.003	0.095	0.005
Maximum	0.529	< 0.001	0.639	0.169	0.153
Median	0.135	0.045	0.067	0.71	0.055
Mean	0.007	< 0.001	0.686	0.917	0.652
Variance	0.174	< 0.001	0.542	0.346	0.495
Standard Deviation	0.147	< 0.001	0.512	0.149	0.179
Percentile Low Risk	0.024	0.853	0.719	0.786	0.541
Percentile Medium Risk	0.266	0.366	0.625	0.452	0.433
Percentile High Risk	0.16	0.235	0.059	0.029	< 0.001
Percentile Very High Risk	< 0.001	0.584	0.105	0.18	0.15
Alves Low Risk	0.072	< 0.001	0.085	0.039	0.01
Alves Medium Risk	0.405	0.116	0.134	0.027	0.048
Alves High Risk	0.095	< 0.001	0.606	0.348	0.325
Alves Very High Risk	0.002	0.006	0.076	0.286	0.261
Type Lines of Code (TLOC)					
Minimum	0.653	< 0.001	0.133	0.019	0.005
Maximum	0.39	0.26	0.621	0.886	0.354
Median	0.017	0.216	0.262	0.192	0.406
Mean	0.003	0.753	0.376	0.409	0.507
Variance	0.13	0.052	0.177	0.109	0.091
Standard Deviation	0.207	0.064	0.187	0.164	0.082
Percentile Low Risk	< 0.001	0.85	0.589	0.665	0.839
Percentile Medium Risk	0.006	0.99	0.036	0.046	0.025
Percentile High Risk	0.035	0.466	0.193	0.024	0.028
Percentile Very High Risk	< 0.001	0.961	0.08	0.079	0.265
Alves Low Risk	0.067	0.204	0.371	0.233	0.33
Alves Medium Risk	0.436	0.025	0.287	0.123	0.293
Alves High Risk	0.642	0.974	0.586	0.529	0.537
Alves Very High Risk	0.081	0.773	0.759	0.158	0.247

Variable	Group P-value	Project P-value	TA's (A) P-value	TA's (B) P-value	TA's (C) P-value
Weighted Methods per Class (WMC)					
Minimum	0.332	0.044	0.212	0.385	0.285
Maximum	0.315	0.014	0.496	0.905	0.283
Median	0.189	0.012	0.389	0.669	0.73
Mean	0.003	0.668	0.491	0.926	0.696
Variance	0.019	0.574	0.049	0.49	0.142
Standard Deviation	0.057	0.496	0.099	0.681	0.146
Percentile Low Risk	0.003	0.638	0.993	0.979	0.996
Percentile Medium Risk	0.133	0.042	0.668	0.491	0.69
Percentile High Risk	0.18	0.399	0.724	0.861	0.799
Percentile Very High Risk	< 0.001	0.867	0.056	0.058	0.142
Alves Low Risk	0.127	0.617	0.535	0.42	0.49
Alves Medium Risk	0.52	0.966	0.432	0.174	0.279
Alves High Risk	0.86	0.088	0.62	0.738	0.443
Alves Very High Risk	0.027	0.385	0.401	0.315	0.3
Number of Children (NC)					
Minimum	-	-	-	-	-
Maximum	0.49	0.006	0.021	0.065	0.058
Median	-	-	-	-	-
Mean	0.973	< 0.001	0.532	0.587	0.622
Variance	0.783	< 0.001	0.184	0.191	0.454
Standard Deviation	0.773	< 0.001	0.07	0.16	0.134
Percentile Low Risk	-	-	-	-	-
Percentile Medium Risk	-	-	-	-	-
Percentile High Risk	1	< 0.001	0.966	0.019	0.051
Percentile Very High Risk	1	< 0.001	0.966	0.019	0.051
Alves Low Risk	0.499	< 0.001	0.224	0.016	0.078
Alves Medium Risk	0.707	0.099	0.048	0.001	0.008
Alves High Risk	0.913	0.414	0.637	0.32	0.272
Alves Very High Risk	0.981	< 0.001	0.828	0.846	0.942
Depth of Inheritance Tree (DIT)					
Minimum	-	-	-	-	-
Maximum	0.796	< 0.001	0.067	0.292	0.056
Median	0.942	< 0.001	0.712	0.733	0.897
Mean	0.971	< 0.001	0.544	0.635	0.647
Variance	0.941	< 0.001	0.542	0.653	0.652
Standard Deviation	0.864	< 0.001	0.144	0.424	0.138
Percentile Low Risk	1	< 0.001	0.778	0.006	0.02
Percentile Medium Risk	-	-	-	-	-
Percentile High Risk	0.996	< 0.001	0.83	0.565	0.721
Percentile Very High Risk	1	< 0.001	0.917	0.01	0.029
Alves Low Risk	0.037	0.044	0.02	0.164	0.047
Alves Medium Risk	0.416	0.122	0.084	0.01	0.014
Alves High Risk	0.079	0.41	0.15	0.595	0.404
Alves Very High Risk	0.652	0.019	0.181	0.396	0.377

Variable	Group P-value	Project P-value	TA's (A) P-value	TA's (B) P-value	TA's (C) P-value
Lack of Cohesion Of Methods (LCOM)					
Minimum	0.522	0.608	0.438	0.052	0.211
Maximum	0.971	< 0.001	0.679	0.529	0.281
Median	0.219	0.238	0.559	0.508	0.82
Mean	0.324	< 0.001	0.5	0.347	0.643
Variance	0.818	< 0.001	0.598	0.574	0.538
Standard Deviation	0.867	< 0.001	0.537	0.635	0.504
Percentile Low Risk	0.999	< 0.001	0.874	0.154	0.175
Percentile Medium Risk	1	< 0.001	0.353	0.025	0.041
Percentile High Risk	1	< 0.001	0.987	0.261	0.229
Percentile Very High Risk	0.243	0.064	0.408	0.424	0.681
Alves Low Risk	0.103	< 0.001	0.176	0.057	0.12
Alves Medium Risk	0.006	0.067	0.058	0.097	0.021
Alves High Risk	0.573	< 0.001	0.143	0.008	0.016
Alves Very High Risk	0.624	< 0.001	0.291	0.349	0.526
Fan-in (FANIN)					
Minimum	0.489	0.37	0.144	0.064	< 0.001
Maximum	0.619	< 0.001	0.231	0.034	0.048
Median	0.036	< 0.001	0.159	0.099	0.099
Mean	0.137	< 0.001	0.797	0.963	0.932
Variance	0.828	< 0.001	0.312	0.046	0.072
Standard Deviation	0.576	< 0.001	0.176	0.035	0.044
Percentile Low Risk	0.087	< 0.001	0.189	0.096	0.151
Percentile Medium Risk	0.998	< 0.001	0.49	0.738	0.785
Percentile High Risk	0.894	< 0.001	0.2	0.016	0.006
Percentile Very High Risk	0.359	< 0.001	0.473	0.931	0.429
Alves Low Risk	0.174	< 0.001	0.386	0.094	0.031
Alves Medium Risk	0.05	0.342	0.283	0.51	0.228
Alves High Risk	0.143	< 0.001	0.403	< 0.001	0.003
Alves Very High Risk	0.875	< 0.001	0.002	0.1	0.008
Fan-out (FANOUT)					
Minimum	0.659	0.026	0.966	0.934	0.455
Maximum	0.133	0.013	0.767	0.757	0.759
Median	0.237	< 0.001	0.601	0.618	0.817
Mean	0.137	< 0.001	0.797	0.963	0.932
Variance	0.033	< 0.001	0.871	0.96	0.933
Standard Deviation	0.051	< 0.001	0.721	0.904	0.703
Percentile Low Risk	0.929	< 0.001	0.945	0.985	0.649
Percentile Medium Risk	1	< 0.001	0.307	0.925	0.451
Percentile High Risk	0.99	< 0.001	0.696	0.307	0.46
Percentile Very High Risk	0.043	0.015	0.779	0.551	0.435
Alves Low Risk	0.135	< 0.001	0.453	0.236	0.102
Alves Medium Risk	0.883	0.003	0.657	0.365	0.149
Alves High Risk	0.221	0.02	0.068	0.099	0.015
Alves Very High Risk	0.035	0.156	0.929	0.841	0.904

4.3 TukeyHSD Result Tables

In the following tables, each of the four categories Project, and categorizations A,B and C for Teaching assistants are shown. Each of these receives its own table, one for all data, and one for complete data. In them, the category the variable falls under, the aggregation used for it, the p-value, mean difference, minimal difference and maximal difference will be shown, within a 95% confidence interval. This is repeated for each difference.

Just like with the ANOVA tables, each result is given a coloured cell as follows:

Significance level Assumption	[1,0.1)	[0.1, 0.05)	[0.05, 0.01)	[0.01, 0.001)	[0.001, 0)
All Assumptions Met					
Homogeneous, not Normal					
No Assumptions Met					

These labels are intended to distinguish between different levels of significance, as well as differentiating between which assumptions the ANOVA the TukeyHSD was based on are met according to the Shapiro-Wilk normality test (indicated as 'Normal' above) and Levene's test for homogeneity of variance ('Homogeneous') To this end a colour is added to each cell to indicate the assumptions met, the opacity of the colour relates to the level of significance.

The astute reader will notice that there is no colour option for Normal, not Homogeneous. To arrive at this approach, some tests with the data were performed, and the results of those tests showed that such a situation does not occur in the data used in this thesis.

Table 4.3: Tukey differences for Projects on all data

	graphEditor-cardGame		**rpg-cardGame**		**rpg-graphEditor**	
	p-value	difference	p-value	difference	p-value	difference
	lower bound	upper bound	lower bound	upper bound	lower bound	upper bound
Resultant Variable						
Coding Standard			< 0.001	-0.197		
			-0.298	-0.097		
Total						
Number of Methods	0.001	26.23	< 0.001	-39.68	< 0.001	-65.91
	8.818	43.642	-55.345	-24.016	-83.042	-48.779
Number of Types	0.005	4.853	< 0.001	-6.451	< 0.001	-11.304
	1.235	8.471	-9.706	-3.196	-14.864	-7.744
Method Lines of Code (MLOC)						
Minimum	< 0.001	1.457	0.095	-0.284	< 0.001	-1.741
	1.099	1.815	-0.606	0.038	-2.093	-1.389
Maximum	0.306	15.379	< 0.001	68.591	< 0.001	53.212
	-9.281	40.04	46.405	90.776	28.948	77.475
Median	< 0.001	0.518	0.189	-0.211	< 0.001	-0.729
	0.203	0.834	-0.495	0.073	-1.04	-0.419
Mean	0.03	0.807	< 0.001	2.188	< 0.001	1.381
	0.064	1.551	1.519	2.857	0.65	2.113
Variance	0.987	11.852	< 0.001	320.526	< 0.001	308.674
	-171.733	195.437	155.369	485.684	128.046	489.303
Standard Deviation	0.509	1.492	< 0.001	9.775	< 0.001	8.284
	-1.683	4.666	6.919	12.631	5.16	11.407
Percentile Low Risk	0.031	-0.042	0.081	-0.032	0.812	0.01
	-0.081	-0.003	-0.067	0.003	-0.028	0.048
Percentile Medium Risk	< 0.001	0.046	< 0.001	0.055	0.522	0.01
	0.024	0.067	0.036	0.075	-0.011	0.031
Alves Low Risk	0.034	0.017	< 0.001	0.046	< 0.001	0.029
	0.001	0.033	0.031	0.06	0.013	0.044
Alves Medium Risk	0.109	-0.009	< 0.001	-0.022	0.007	-0.013
	-0.019	0.001	-0.031	-0.013	-0.023	-0.003
Alves High Risk	0.337	-0.005	< 0.001	-0.017	0.001	-0.012
	-0.013	0.003	-0.025	-0.01	-0.02	-0.004
Alves Very High Risk	0.354	-0.003	0.004	-0.006	0.279	-0.003
	-0.008	0.002	-0.011	-0.002	-0.008	0.002

	graphEditor-cardGame		rpg-cardGame		rpg-graphEditor	
	p-value lower bound	difference upper bound	p-value lower bound	difference upper bound	p-value lower bound	difference upper bound
Cyclomatic Complexity (CC)						
Maximum	1 -2.911	-0.036 2.838	< 0.001 5.547	8.133 10.719	< 0.001 5.341	8.169 10.997
Mean	0.932 -0.123	-0.016 0.09	< 0.001 0.153	0.249 0.344	< 0.001 0.16	0.265 0.37
Variance	0.98 -2.927	-0.218 2.491	< 0.001 2.63	5.067 7.504	< 0.001 2.62	5.285 7.95
Standard Deviation	0.959 -0.415	-0.043 0.328	< 0.001 0.783	1.117 1.451	< 0.001 0.795	1.16 1.525
Percentile High Risk	< 0.001 -0.056	-0.035 -0.013	0.006 -0.045	-0.026 -0.006	0.593 -0.012	0.009 0.03
Percentile Very High Risk	0.181 -0.006	0.017 0.039	0.006 0.006	0.027 0.047	0.549 -0.012	0.01 0.032
Alves Low Risk	0.838 -0.012	0.004 0.02	< 0.001 0.044	0.059 0.073	< 0.001 0.039	0.055 0.071
Alves Medium Risk	0.902 -0.013	-0.002 0.009	< 0.001 -0.036	-0.026 -0.017	< 0.001 -0.035	-0.024 -0.014
Alves High Risk	0.959 -0.01	-0.001 0.008	< 0.001 -0.032	-0.024 -0.016	< 0.001 -0.032	-0.023 -0.015
Alves Very High Risk	0.924 -0.007	-0.001 0.005	< 0.001 -0.014	-0.008 -0.003	0.01 -0.013	-0.007 -0.001
Parameter Count (PC)						
Maximum	< 0.001 1.712	2.391 3.069	0.009 0.162	0.772 1.382	< 0.001 -2.286	-1.619 -0.951
Median	< 0.001 0.623	0.741 0.859	< 0.001 0.629	0.735 0.841	0.991 -0.123	-0.006 0.11
Mean	< 0.001 0.25	0.33 0.41	< 0.001 0.28	0.352 0.424	0.785 -0.057	0.022 0.101
Variance	< 0.001 0.203	0.428 0.652	< 0.001 0.176	0.377 0.579	0.852 -0.271	-0.05 0.17
Standard Deviation	< 0.001 0.148	0.249 0.351	< 0.001 0.087	0.178 0.27	0.216 -0.171	-0.071 0.029
Percentile Low Risk	< 0.001 -0.214	-0.172 -0.131	< 0.001 -0.245	-0.207 -0.17	0.109 -0.076	-0.035 0.006
Percentile High Risk	< 0.001 0.417	0.452 0.488	< 0.001 0.479	0.511 0.543	< 0.001 0.024	0.059 0.094
Percentile Very High Risk	< 0.001 -0.318	-0.28 -0.243	< 0.001 -0.338	-0.304 -0.27	0.277 -0.061	-0.024 0.013
Alves Low Risk	0.911 -0.038	-0.006 0.026	0.004 -0.068	-0.039 -0.01	0.033 -0.065	-0.034 -0.002
Alves Very High Risk	0.553 -0.033	-0.01 0.013	< 0.001 0.016	0.037 0.057	< 0.001 0.024	0.047 0.069

	graphEditor-cardGame		rpg-cardGame		rpg-graphEditor	
	p-value lower bound	difference upper bound	p-value lower bound	difference upper bound	p-value lower bound	difference upper bound
Number Of Fields (NOF)						
Median	0.038	0.496	0.242	0.292	0.557	-0.204
	0.022	0.971	-0.135	0.719	-0.671	0.263
Percentile Low Risk	< 0.001	-0.159	0.044	-0.069	0.012	0.09
	-0.234	-0.084	-0.137	-0.001	0.016	0.164
Percentile Medium Risk	< 0.001	0.201	< 0.001	0.072	< 0.001	-0.129
	0.159	0.243	0.034	0.11	-0.171	-0.087
Alves Low Risk	0.979	0.004	0.009	-0.055	0.011	-0.059
	-0.045	0.053	-0.099	-0.012	-0.107	-0.011
Alves Very High Risk	0.962	-0.003	0.071	0.021	0.059	0.024
	-0.028	0.022	-0.001	0.044	-0.001	0.049
Number Of Private Fields (NOPF)						
Alves Very High Risk	0.883	-0.01	0.072	0.043	0.035	0.053
	-0.061	0.041	-0.003	0.088	0.003	0.103
Number Of Methods (NOM)						
Maximum	< 0.001	9.143	0.002	-5.162	< 0.001	-14.305
	5.277	13.01	-8.64	-1.683	-18.11	-10.501
Median	0.05	0.554	0.179	0.376	0.72	-0.179
	0	1.109	-0.123	0.874	-0.724	0.367
Mean	0.309	0.316	0.013	-0.552	< 0.001	-0.868
	-0.192	0.824	-1.009	-0.094	-1.368	-0.368
Variance	0.018	10.01	< 0.001	-12.18	< 0.001	-22.19
	1.418	18.603	-19.91	-4.45	-30.644	-13.736
Standard Deviation	0.044	0.73	< 0.001	-1.421	< 0.001	-2.152
	0.015	1.446	-2.065	-0.778	-2.856	-1.448
Percentile Low Risk	0.72	-0.022	0.191	0.046	0.047	0.068
	-0.091	0.046	-0.016	0.107	0.001	0.135
Percentile Medium Risk	0.67	0.017	0.051	-0.042	0.008	-0.059
	-0.03	0.063	-0.084	0	-0.105	-0.013
Alves Low Risk	0.663	0.011	0.022	-0.03	0.003	-0.041
	-0.019	0.041	-0.057	-0.004	-0.071	-0.012
Alves Very High Risk	0.722	-0.005	0.193	0.01	0.048	0.015
	-0.02	0.01	-0.004	0.024	0	0.03

	graphEditor-cardGame		rpg-cardGame		rpg-graphEditor	
	p-value	difference	p-value	difference	p-value	difference
	lower bound	upper bound	lower bound	upper bound	lower bound	upper bound
Number Of Private Methods (NOPM)						
Minimum	< 0.001	0.625	< 0.001	0.314	0.002	-0.311
	0.41	0.839	0.121	0.506	-0.522	-0.1
Maximum	< 0.001	10.5	0.01	-4.021	< 0.001	-14.52
	6.917	14.082	-7.244	-0.798	-18.045	-10.995
Mean	0.015	0.572	0.346	-0.255	< 0.001	-0.827
	0.092	1.052	-0.687	0.177	-1.3	-0.355
Variance	< 0.001	14.969	0.037	-7.144	< 0.001	-22.113
	7.414	22.525	-13.941	-0.346	-29.547	-14.679
Standard Deviation	< 0.001	1.233	< 0.001	-1.023	< 0.001	-2.255
	0.566	1.9	-1.623	-0.423	-2.912	-1.599
Alves Low Risk	0.098	0.028	0.002	-0.042	< 0.001	-0.07
	-0.004	0.059	-0.07	-0.014	-0.101	-0.039
Alves High Risk	0.336	-0.012	0.06	0.017	0.001	0.029
	-0.032	0.008	-0.001	0.035	0.01	0.048
Alves Very High Risk	0.506	-0.007	0.018	0.016	0.001	0.023
	-0.023	0.008	0.002	0.03	0.008	0.039
Type Lines of Code (TLOC)						
Minimum	0.009	1.794	< 0.001	-3.259	< 0.001	-5.053
	0.37	3.218	-4.539	-1.978	-6.453	-3.652
Weighted Methods per Class (WMC)						
Maximum	0.029	9.185	0.424	-4.018	< 0.001	-13.203
	0.762	17.607	-11.595	3.559	-21.49	-4.916
Median	0.049	0.873	0.002	1.15	0.726	0.277
	0.003	1.743	0.367	1.932	-0.579	1.133
Number of Children (NC)						
Maximum	0.152	-0.74	0.242	0.577	0.003	1.316
	-1.677	0.197	-0.266	1.42	0.394	2.239
Mean	0.556	-0.047	< 0.001	0.295	< 0.001	0.342
	-0.155	0.06	0.198	0.391	0.236	0.448
Variance	0.942	0.064	< 0.001	0.73	0.002	0.666
	-0.396	0.524	0.316	1.144	0.213	1.119
Standard Deviation	0.142	-0.174	< 0.001	0.307	< 0.001	0.481
	-0.39	0.042	0.112	0.502	0.268	0.694
Percentile High Risk	1	0	< 0.001	0.878	< 0.001	0.878
	-0.033	0.033	0.848	0.907	0.845	0.91
Percentile Very High Risk	1	0	< 0.001	-0.878	< 0.001	-0.878
	-0.033	0.033	-0.907	-0.848	-0.91	-0.845
Alves Low Risk	0.378	-0.029	< 0.001	-0.097	0.004	-0.069
	-0.08	0.022	-0.144	-0.051	-0.119	-0.018
Alves Medium Risk	0.807	-0.008	0.001	-0.041	0.024	-0.033
	-0.037	0.022	-0.067	-0.014	-0.062	-0.004
Alves Very High Risk	0.335	0.028	< 0.001	0.135	< 0.001	0.107
	-0.019	0.074	0.093	0.177	0.061	0.153

	graphEditor-cardGame		rpg-cardGame		rpg-graphEditor	
	p-value	difference	p-value	difference	p-value	difference
	lower bound	upper bound	lower bound	upper bound	lower bound	upper bound
Depth of Inheritance Tree (DIT)						
Maximum	< 0.001	-0.631	0.047	0.326	< 0.001	0.957
	-0.989	-0.274	0.004	0.648	0.605	1.309
Median	0.99	0.009	< 0.001	0.456	< 0.001	0.446
	-0.158	0.177	0.305	0.607	0.281	0.611
Mean	0.697	-0.049	< 0.001	0.386	< 0.001	0.435
	-0.191	0.093	0.258	0.514	0.295	0.574
Variance	0.614	-0.052	< 0.001	0.316	< 0.001	0.367
	-0.181	0.078	0.199	0.432	0.24	0.494
Standard Deviation	0.004	-0.173	< 0.001	0.198	< 0.001	0.371
	-0.3	-0.047	0.084	0.311	0.247	0.495
Percentile Low Risk	1	0	< 0.001	0.694	< 0.001	0.694
	-0.077	0.077	0.625	0.763	0.618	0.769
Percentile High Risk	1	0	< 0.001	0.164	< 0.001	0.164
	-0.044	0.044	0.125	0.204	0.121	0.207
Percentile Very High Risk	1	0	< 0.001	-0.858	< 0.001	-0.858
	-0.052	0.052	-0.905	-0.811	-0.909	-0.807
Alves Low Risk	0.048	0.082	0.94	0.01	0.09	-0.072
	0.001	0.164	-0.063	0.084	-0.152	0.009
Alves Very High Risk	0.053	-0.058	0.965	0.006	0.026	0.064
	-0.117	0	-0.047	0.059	0.006	0.122
Lack of Cohesion Of Methods (LCOM)						
Maximum	< 0.001	0.179	< 0.001	0.425	< 0.001	0.246
	0.067	0.291	0.324	0.525	0.135	0.356
Mean	< 0.001	0.133	< 0.001	0.232	0.017	0.099
	0.047	0.219	0.154	0.309	0.014	0.183
Variance	0.235	-0.033	< 0.001	0.134	< 0.001	0.167
	-0.081	0.015	0.091	0.177	0.12	0.214
Standard Deviation	0.068	-0.04	< 0.001	0.107	< 0.001	0.147
	-0.082	0.002	0.069	0.145	0.106	0.189
Percentile Low Risk	< 0.001	-0.107	< 0.001	0.297	< 0.001	0.403
	-0.17	-0.044	0.24	0.354	0.341	0.466
Percentile Medium Risk	1	0	< 0.001	0.114	< 0.001	0.114
	-0.028	0.028	0.089	0.139	0.086	0.142
Percentile High Risk	< 0.001	0.113	< 0.001	-0.44	< 0.001	-0.553
	0.061	0.165	-0.487	-0.393	-0.604	-0.501
Alves Low Risk	0.377	-0.033	< 0.001	-0.16	< 0.001	-0.126
	-0.092	0.026	-0.213	-0.107	-0.184	-0.068
Alves High Risk	0.226	0.024	< 0.001	0.077	0.001	0.052
	-0.01	0.059	0.045	0.108	0.018	0.087
Alves Very High Risk	0.822	0.011	< 0.001	0.07	0.005	0.058
	-0.033	0.055	0.03	0.109	0.015	0.102

	graphEditor-cardGame		rpg-cardGame		rpg-graphEditor	
	p-value lower bound	difference upper bound	p-value lower bound	difference upper bound	p-value lower bound	difference upper bound
Fan-in (FANIN)						
Maximum	< 0.001 4.982	7.821 10.66	0.401 -3.952	-1.398 1.156	< 0.001 -12.012	-9.219 -6.426
Median	0.983 -0.279	-0.019 0.241	< 0.001 0.214	0.448 0.682	< 0.001 0.212	0.467 0.723
Mean	< 0.001 0.377	0.571 0.764	< 0.001 0.253	0.427 0.601	0.178 -0.334	-0.144 0.047
Variance	< 0.001 8.302	11.124 13.945	0.882 -2.025	0.514 3.052	< 0.001 -13.386	-10.61 -7.834
Standard Deviation	< 0.001 1.043	1.478 1.912	0.711 -0.26	0.13 0.521	< 0.001 -1.775	-1.347 -0.92
Percentile Low Risk	0.767 -0.054	0.022 0.099	< 0.001 -0.189	-0.12 -0.051	< 0.001 -0.218	-0.142 -0.067
Percentile Medium Risk	< 0.001 0.128	0.18 0.232	< 0.001 0.244	0.291 0.338	< 0.001 0.06	0.111 0.163
Percentile High Risk	< 0.001 -0.175	-0.124 -0.073	0.001 -0.116	-0.07 -0.024	0.031 0.004	0.054 0.104
Percentile Very High Risk	< 0.001 -0.119	-0.079 -0.038	< 0.001 -0.138	-0.102 -0.065	0.358 -0.063	-0.023 0.017
Alves Low Risk	0.55 -0.023	0.018 0.059	< 0.001 -0.108	-0.071 -0.034	< 0.001 -0.13	-0.09 -0.049
Alves High Risk	0.429 -0.034	-0.012 0.01	0.014 0.004	0.024 0.044	< 0.001 0.014	0.035 0.057
Alves Very High Risk	0.351 -0.018	-0.007 0.005	< 0.001 0.013	0.023 0.033	< 0.001 0.018	0.03 0.041

	graphEditor-cardGame		rpg-cardGame		rpg-graphEditor	
	p-value	difference	p-value	difference	p-value	difference
	lower bound	upper bound	lower bound	upper bound	lower bound	upper bound
Fan-out (FANOUT)						
Minimum	0.011	0.094	0.659	0.025	0.079	-0.069
	0.018	0.171	-0.043	0.094	-0.144	0.006
Maximum	0.373	0.593	< 0.001	1.569	0.066	0.976
	-0.449	1.635	0.632	2.507	-0.049	2.002
Median	< 0.001	0.459	0.019	0.231	0.039	-0.228
	0.237	0.681	0.031	0.431	-0.446	-0.01
Mean	< 0.001	0.571	< 0.001	0.427	0.178	-0.144
	0.377	0.764	0.253	0.601	-0.334	0.047
Variance	1	0.003	< 0.001	1.061	< 0.001	1.058
	-0.659	0.665	0.466	1.657	0.407	1.709
Standard Deviation	0.884	-0.042	< 0.001	0.307	< 0.001	0.349
	-0.251	0.167	0.119	0.495	0.144	0.554
Percentile Low Risk	< 0.001	0.145	< 0.001	0.253	< 0.001	0.109
	0.078	0.211	0.194	0.313	0.043	0.174
Percentile Medium Risk	< 0.001	-0.377	< 0.001	-0.377	1	0
	-0.414	-0.341	-0.41	-0.345	-0.036	0.036
Percentile High Risk	< 0.001	0.169	< 0.001	0.105	0.012	-0.064
	0.116	0.222	0.057	0.153	-0.116	-0.012
Percentile Very High Risk	0.002	0.064	0.491	0.019	0.045	-0.045
	0.019	0.109	-0.021	0.059	-0.089	-0.001
Alves Low Risk	0.016	-0.043	0.027	0.036	< 0.001	0.079
	-0.08	-0.006	0.003	0.069	0.043	0.115
Alves Medium Risk	0.202	0.015	0.019	-0.021	< 0.001	-0.036
	-0.005	0.035	-0.039	-0.003	-0.055	-0.016
Alves High Risk	0.031	0.023	0.561	-0.008	0.002	-0.031
	0.002	0.044	-0.027	0.011	-0.052	-0.01

Table 4.4: Tukey differences for Projects on complete data

	graphEditor-cardGame		rpg-cardGame		rpg-graphEditor	
	p-value	difference	p-value	difference	p-value	difference
	lower bound	upper bound	lower bound	upper bound	lower bound	upper bound
Resultant Variable						
Coding Standard			< 0.001	-0.221		
			-0.34	-0.103		
Total						
Number of Methods	0.104	18.102	< 0.001	-45.429	< 0.001	-63.531
	-2.791	38.995	-66.322	-24.535	-84.424	-42.638
Number of Types	0.184	3.163	< 0.001	-7.816	< 0.001	-10.98
	-1.072	7.399	-12.052	-3.581	-15.215	-6.744
Method Lines of Code (MLOC)						
Minimum	< 0.001	1.592	0.212	-0.265	< 0.001	-1.857
	1.22	1.963	-0.637	0.106	-2.229	-1.486
Maximum	0.176	19.776	< 0.001	86.265	< 0.001	66.49
	-6.376	45.927	60.114	112.416	40.339	92.641
Median	0.002	0.541	0.464	-0.184	< 0.001	-0.724
	0.174	0.908	-0.551	0.183	-1.091	-0.358
Mean	0.026	0.896	< 0.001	2.51	< 0.001	1.614
	0.085	1.707	1.699	3.321	0.803	2.425
Variance	0.946	28.872	< 0.001	403.86	< 0.001	374.988
	-187.247	244.99	187.741	619.978	158.87	591.106
Standard Deviation	0.409	1.852	< 0.001	11.609	< 0.001	9.756
	-1.576	5.281	8.18	15.037	6.328	13.185
Percentile Low Risk	0.025	-0.047	0.108	-0.036	0.814	0.011
	-0.089	-0.005	-0.078	0.006	-0.031	0.053
Percentile Medium Risk	< 0.001	0.044	< 0.001	0.059	0.294	0.015
	0.021	0.067	0.035	0.082	-0.009	0.038
Alves Low Risk	0.076	0.015	< 0.001	0.041	< 0.001	0.026
	-0.001	0.031	0.025	0.057	0.01	0.042
Alves Medium Risk	0.275	-0.007	< 0.001	-0.018	0.036	-0.011
	-0.018	0.004	-0.029	-0.008	-0.022	-0.001
Alves High Risk	0.466	-0.005	< 0.001	-0.017	0.006	-0.012
	-0.014	0.005	-0.026	-0.008	-0.021	-0.003

	graphEditor-cardGame		**rpg-cardGame**		**rpg-graphEditor**	
	p-value lower bound	difference upper bound	p-value lower bound	difference upper bound	p-value lower bound	difference upper bound
Cyclomatic Complexity (CC)						
Maximum	0.922	-0.531	< 0.001	9.367	< 0.001	9.898
	-3.81	2.749	6.088	12.647	6.619	13.177
Mean	0.975	-0.011	< 0.001	0.285	< 0.001	0.296
	-0.132	0.11	0.164	0.406	0.175	0.417
Variance	0.978	-0.279	< 0.001	6.141	< 0.001	6.42
	-3.591	3.033	2.828	9.453	3.107	9.732
Standard Deviation	0.922	-0.069	< 0.001	1.269	< 0.001	1.338
	-0.494	0.356	0.844	1.694	0.913	1.763
Percentile High Risk	0.016	-0.029	0.406	-0.013	0.29	0.016
	-0.053	-0.004	-0.038	0.011	-0.009	0.04
Percentile Very High Risk	0.186	0.018	0.047	0.024	0.806	0.006
	-0.006	0.042	0	0.048	-0.018	0.03
Alves Low Risk	0.861	0.004	< 0.001	0.059	< 0.001	0.055
	-0.014	0.022	0.041	0.077	0.037	0.073
Alves Medium Risk	0.964	-0.001	< 0.001	-0.026	< 0.001	-0.025
	-0.013	0.011	-0.038	-0.014	-0.037	-0.013
Alves High Risk	0.971	-0.001	< 0.001	-0.024	< 0.001	-0.023
	-0.011	0.009	-0.034	-0.014	-0.033	-0.013
Alves Very High Risk	0.836	-0.002	0.009	-0.009	0.042	-0.007
	-0.009	0.005	-0.016	-0.002	-0.015	0
Parameter Count (PC)						
Maximum	< 0.001	2.612	0.002	1.082	< 0.001	-1.531
	1.867	3.357	0.336	1.827	-2.276	-0.785
Median	< 0.001	0.755	< 0.001	0.755	1	0
	0.629	0.881	0.629	0.881	-0.126	0.126
Mean	< 0.001	0.34	< 0.001	0.36	0.843	0.019
	0.258	0.423	0.277	0.442	-0.063	0.102
Variance	< 0.001	0.449	< 0.001	0.392	0.812	-0.057
	0.23	0.668	0.173	0.611	-0.276	0.162
Standard Deviation	< 0.001	0.261	< 0.001	0.196	0.305	-0.065
	0.157	0.365	0.092	0.3	-0.169	0.039
Percentile Low Risk	< 0.001	-0.178	< 0.001	-0.212	0.168	-0.034
	-0.222	-0.133	-0.257	-0.167	-0.079	0.01
Percentile High Risk	< 0.001	0.455	< 0.001	0.514	0.001	0.059
	0.416	0.493	0.475	0.553	0.021	0.098
Percentile Very High Risk	< 0.001	-0.277	< 0.001	-0.302	0.28	-0.025
	-0.316	-0.239	-0.34	-0.263	-0.063	0.014
Alves Low Risk	0.687	-0.012	0.002	-0.05	0.022	-0.038
	-0.045	0.022	-0.083	-0.016	-0.072	-0.004
Alves High Risk	0.035	0.02	0.581	0.008	0.29	-0.012
	0.001	0.038	-0.011	0.026	-0.03	0.007
Alves Very High Risk	0.54	-0.01	< 0.001	0.035	< 0.001	0.044
	-0.031	0.012	0.013	0.056	0.023	0.066

	graphEditor-cardGame		rpg-cardGame		rpg-graphEditor	
	p-value	difference	p-value	difference	p-value	difference
	lower bound	upper bound	lower bound	upper bound	lower bound	upper bound
Number Of Fields (NOF)						
Median	0.027	0.561	0.242	0.347	0.579	-0.214
	0.053	1.07	-0.162	0.856	-0.723	0.294
Percentile Low Risk	< 0.001	-0.169	0.064	-0.08	0.033	0.089
	-0.252	-0.086	-0.163	0.004	0.006	0.172
Percentile Medium Risk	< 0.001	0.195	0.013	0.056	< 0.001	-0.139
	0.149	0.241	0.01	0.102	-0.185	-0.093
Alves Low Risk	0.995	-0.002	0.008	-0.064	0.011	-0.062
	-0.052	0.048	-0.114	-0.014	-0.112	-0.012
Number Of Private Fields (NOPF)						
Alves High Risk	0.119	-0.031	0.029	-0.04	0.826	-0.009
	-0.068	0.006	-0.077	-0.003	-0.046	0.028
Number Of Methods (NOM)						
Maximum	< 0.001	8.653	0.013	-5.816	< 0.001	-14.469
	3.834	13.472	-10.635	-0.997	-19.289	-9.65
Median	0.036	0.592	0.378	0.316	0.478	-0.276
	0.03	1.153	-0.245	0.878	-0.837	0.286
Mean	0.304	0.332	0.039	-0.554	< 0.001	-0.885
	-0.2	0.863	-1.085	-0.022	-1.417	-0.354
Variance	0.103	9.374	0.01	-13.485	< 0.001	-22.859
	-1.425	20.173	-24.284	-2.686	-33.658	-12.06
Standard Deviation	0.146	0.675	< 0.001	-1.483	< 0.001	-2.158
	-0.172	1.523	-2.331	-0.635	-3.006	-1.31
Alves Low Risk	0.915	0.005	0.002	-0.04	< 0.001	-0.045
	-0.022	0.031	-0.067	-0.013	-0.071	-0.018
Alves Medium Risk	0.704	0.007	0.03	0.024	0.184	0.016
	-0.015	0.029	0.002	0.046	-0.006	0.038
Alves High Risk	0.788	-0.004	0.162	0.012	0.037	0.017
	-0.02	0.011	-0.004	0.028	0.001	0.033

	graphEditor-cardGame		rpg-cardGame		rpg-graphEditor	
	p-value	difference	p-value	difference	p-value	difference
	lower bound	upper bound	lower bound	upper bound	lower bound	upper bound
Number Of Private Methods (NOPM)						
Minimum	< 0.001	0.673	0.003	0.327	0.001	-0.347
	0.443	0.904	0.096	0.557	-0.578	-0.116
Maximum	< 0.001	9.959	0.031	-4.857	< 0.001	-14.816
	5.462	14.456	-9.354	-0.36	-19.313	-10.32
Median	0.042	0.531	0.18	0.388	0.79	-0.143
	0.014	1.047	-0.128	0.904	-0.659	0.373
Mean	0.022	0.582	0.5	-0.245	< 0.001	-0.826
	0.067	1.096	-0.759	0.27	-1.341	-0.312
Variance	0.001	14.21	0.069	-8.911	< 0.001	-23.121
	4.753	23.667	-18.368	0.546	-32.578	-13.664
Standard Deviation	0.002	1.144	0.002	-1.154	< 0.001	-2.298
	0.362	1.927	-1.936	-0.371	-3.08	-1.516
Alves Low Risk	0.125	0.026	0.002	-0.045	< 0.001	-0.071
	-0.005	0.057	-0.076	-0.014	-0.102	-0.04
Alves High Risk	0.564	-0.007	0.004	0.023	< 0.001	0.03
	-0.023	0.009	0.006	0.039	0.013	0.046
Alves Very High Risk	0.244	-0.01	0.238	0.01	0.004	0.019
	-0.024	0.005	-0.004	0.024	0.005	0.034
Type Lines of Code (TLOC)						
Minimum	0.012	1.98	< 0.001	-3.122	< 0.001	-5.102
	0.37	3.589	-4.732	-1.513	-6.712	-3.492
Alves Medium Risk	0.952	0.002	0.07	-0.018	0.034	-0.02
	-0.017	0.022	-0.037	0.001	-0.039	-0.001
Weighted Methods per Class (WMC)						
Minimum	0.038	0.245	0.687	0.082	0.227	-0.163
	0.011	0.479	-0.152	0.316	-0.397	0.071
Maximum	0.134	8.184	0.584	-4.204	0.011	-12.388
	-1.866	18.233	-14.254	5.846	-22.437	-2.338
Median	0.066	0.898	0.013	1.143	0.812	0.245
	-0.045	1.841	0.2	2.086	-0.698	1.188
Percentile Medium Risk	0.547	-0.019	0.303	0.027	0.033	0.046
	-0.062	0.024	-0.016	0.07	0.003	0.089

	graphEditor-cardGame		rpg-cardGame		rpg-graphEditor	
	p-value lower bound	difference upper bound	p-value lower bound	difference upper bound	p-value lower bound	difference upper bound
Number of Children (NC)						
Maximum	0.252	-0.776	0.218	0.816	0.004	1.592
	-1.929	0.378	-0.337	1.969	0.439	2.745
Mean	0.694	-0.041	< 0.001	0.326	< 0.001	0.367
	-0.16	0.078	0.207	0.445	0.248	0.486
Variance	0.946	0.079	0.001	0.909	0.003	0.831
	-0.51	0.667	0.321	1.498	0.242	1.419
Standard	0.265	-0.169	0.003	0.365	< 0.001	0.533
Deviation	-0.424	0.087	0.109	0.62	0.278	0.788
Percentile High	1	0	< 0.001	0.873	< 0.001	0.873
Risk	-0.035	0.035	0.838	0.909	0.838	0.909
Percentile Very	1	0	< 0.001	-0.873	< 0.001	-0.873
High Risk	-0.035	0.035	-0.909	-0.838	-0.909	-0.838
Alves Low Risk	0.251	-0.035	< 0.001	-0.117	< 0.001	-0.082
	-0.088	0.017	-0.17	-0.064	-0.134	-0.029
Alves Very	0.261	0.031	< 0.001	0.137	< 0.001	0.106
High Risk	-0.016	0.078	0.09	0.184	0.059	0.153
Depth of Inheritance Tree (DIT)						
Maximum	< 0.001	-0.633	0.046	0.388	< 0.001	1.02
	-1.015	-0.25	0.005	0.77	0.638	1.403
Median	0.991	0.01	< 0.001	0.48	< 0.001	0.469
	-0.178	0.198	0.292	0.668	0.281	0.657
Mean	0.809	-0.041	< 0.001	0.432	< 0.001	0.473
	-0.198	0.116	0.275	0.588	0.316	0.629
Variance	0.733	-0.044	< 0.001	0.347	< 0.001	0.391
	-0.183	0.095	0.209	0.486	0.253	0.53
Standard	0.01	-0.168	< 0.001	0.221	< 0.001	0.389
Deviation	-0.303	-0.033	0.087	0.356	0.255	0.524
Percentile Low	1	0	< 0.001	0.674	< 0.001	0.674
Risk	-0.083	0.083	0.591	0.757	0.591	0.757
Percentile High	1	0	< 0.001	0.166	< 0.001	0.166
Risk	-0.045	0.045	0.121	0.21	0.121	0.21
Percentile Very	1	0	< 0.001	-0.839	< 0.001	-0.839
High Risk	-0.058	0.058	-0.897	-0.782	-0.897	-0.782
Alves Low Risk	0.076	0.083	1	0	0.077	-0.083
	-0.007	0.174	-0.09	0.09	-0.173	0.007
Alves Very	0.13	-0.055	0.679	0.024	0.017	0.078
High Risk	-0.121	0.012	-0.043	0.09	0.012	0.145

	graphEditor-cardGame		rpg-cardGame		rpg-graphEditor	
	p-value lower bound	difference upper bound	p-value lower bound	difference upper bound	p-value lower bound	difference upper bound
Lack of Cohesion Of Methods (LCOM)						
Maximum	0.013 0.027	0.154 0.281	< 0.001 0.297	0.424 0.551	< 0.001 0.143	0.27 0.397
Mean	0.002 0.046	0.14 0.234	< 0.001 0.156	0.25 0.344	0.017 0.016	0.11 0.204
Variance	0.152 -0.093	-0.041 0.011	< 0.001 0.086	0.138 0.189	< 0.001 0.127	0.178 0.23
Standard Deviation	0.041 -0.094	-0.048 -0.001	< 0.001 0.065	0.111 0.157	< 0.001 0.112	0.159 0.205
Percentile Low Risk	< 0.001 -0.184	-0.114 -0.044	< 0.001 0.208	0.278 0.348	< 0.001 0.322	0.392 0.462
Percentile Medium Risk	1 -0.025	0 0.025	< 0.001 0.087	0.112 0.137	< 0.001 0.087	0.112 0.137
Percentile High Risk	< 0.001 0.057	0.119 0.181	< 0.001 -0.492	-0.429 -0.367	< 0.001 -0.611	-0.549 -0.486
Alves Low Risk	0.586 -0.09	-0.027 0.037	< 0.001 -0.227	-0.163 -0.099	< 0.001 -0.2	-0.136 -0.073
Alves High Risk	0.484 -0.019	0.018 0.055	< 0.001 0.031	0.068 0.105	0.004 0.013	0.05 0.087
Alves Very High Risk	0.751 -0.032	0.014 0.061	< 0.001 0.032	0.079 0.125	0.004 0.018	0.065 0.111
Fan-in (FANIN)						
Maximum	< 0.001 3.815	7.265 10.716	0.404 -5.328	-1.878 1.573	< 0.001 -12.594	-9.143 -5.692
Median	0.996 -0.272	0.01 0.292	0.001 0.147	0.429 0.711	0.002 0.136	0.418 0.7
Mean	< 0.001 0.318	0.535 0.752	< 0.001 0.207	0.424 0.642	0.453 -0.328	-0.11 0.107
Variance	< 0.001 7.194	10.714 14.235	0.971 -3.178	0.342 3.863	< 0.001 -13.893	-10.372 -6.852
Standard Deviation	< 0.001 0.835	1.348 1.862	0.883 -0.41	0.103 0.616	< 0.001 -1.759	-1.245 -0.732
Percentile Low Risk	0.898 -0.07	0.016 0.102	0.003 -0.208	-0.122 -0.036	< 0.001 -0.224	-0.138 -0.053
Percentile Medium Risk	< 0.001 0.119	0.179 0.239	< 0.001 0.234	0.294 0.354	< 0.001 0.055	0.114 0.174
Percentile High Risk	< 0.001 -0.168	-0.115 -0.062	0.005 -0.123	-0.071 -0.018	0.114 -0.008	0.045 0.098
Percentile Very High Risk	< 0.001 -0.122	-0.08 -0.038	< 0.001 -0.143	-0.101 -0.059	0.471 -0.063	-0.021 0.021
Alves Low Risk	0.507 -0.022	0.02 0.061	< 0.001 -0.111	-0.07 -0.028	< 0.001 -0.131	-0.089 -0.048
Alves High Risk	0.388 -0.036	-0.013 0.01	0.021 0.003	0.026 0.05	< 0.001 0.016	0.039 0.062
Alves Very High Risk	0.298 -0.019	-0.007 0.004	< 0.001 0.013	0.024 0.036	< 0.001 0.02	0.031 0.043

	graphEditor-cardGame		rpg-cardGame		rpg-graphEditor	
	p-value	difference	p-value	difference	p-value	difference
	lower bound	upper bound	lower bound	upper bound	lower bound	upper bound
Fan-out (FANOUT)						
Minimum	0.028	0.102	0.863	0.02	0.099	-0.082
	0.009	0.195	-0.073	0.114	-0.175	0.012
Maximum	0.884	0.245	0.016	1.449	0.056	1.204
	-0.983	1.472	0.222	2.676	-0.023	2.432
Median	< 0.001	0.449	0.037	0.255	0.146	-0.194
	0.206	0.692	0.012	0.498	-0.437	0.049
Mean	< 0.001	0.535	< 0.001	0.424	0.453	-0.11
	0.318	0.752	0.207	0.642	-0.328	0.107
Variance	0.968	-0.079	0.003	1.09	0.001	1.169
	-0.848	0.69	0.321	1.859	0.4	1.937
Standard Deviation	0.714	-0.079	0.009	0.304	< 0.001	0.383
	-0.317	0.16	0.065	0.542	0.144	0.621
Percentile Low Risk	< 0.001	0.171	< 0.001	0.257	0.022	0.086
	0.094	0.247	0.18	0.333	0.01	0.162
Percentile Medium Risk	< 0.001	-0.387	< 0.001	-0.387	1	0
	-0.426	-0.349	-0.426	-0.349	-0.039	0.039
Percentile High Risk	< 0.001	0.157	< 0.001	0.119	0.289	-0.038
	0.097	0.216	0.059	0.178	-0.098	0.022
Percentile Very High Risk	0.017	0.06	0.837	0.012	0.071	-0.048
	0.009	0.112	-0.039	0.064	-0.099	0.003
Alves Low Risk	0.151	-0.036	0.1	0.039	< 0.001	0.075
	-0.081	0.009	-0.006	0.085	0.03	0.12
Alves Medium Risk	0.749	0.007	0.026	-0.026	0.003	-0.033
	-0.016	0.031	-0.049	-0.003	-0.056	-0.01
Alves High Risk	0.047	0.026	0.993	-0.001	0.035	-0.027
	0	0.052	-0.027	0.025	-0.053	-0.002

Table 4.5: Tukey differences for TA teams Categorisation A on all data

B-A	**C-A**	**D-A**	**E-A**	**F-A**	**C-B**	**D-B**	**E-B**	**F-B**	**D-C**	**E-C**	**F-C**	**E-D**	**F-D**	**F-E**
pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl
dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif
lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob
upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb
Resultant Variable														
OO Standard														
0.28	0	0.73	0.08	0.8	0.37	0.02	0.92	1	0	0.99	0.41	0.01	0.2	0.86
-0.15	-0.3	0.12	-0.24	-0.12	-0.14	0.28	-0.09	0.03	0.42	0.06	0.18	-0.36	-0.24	0.12
-0.36	-0.51	-0.13	-0.5	-0.39	-0.35	0.03	-0.34	-0.23	0.16	-0.21	-0.09	-0.66	-0.54	-0.18
0.05	-0.08	0.38	0.02	0.15	0.07	0.52	0.17	0.3	0.68	0.32	0.45	-0.07	0.06	0.43
Coding Standard														
0.33	0	0.05	0.95	0.97	0.12	0.81	0.12	0.97	0.95	0	0.07	0.02	0.51	0.72
-0.15	-0.33	-0.25	0.08	-0.07	-0.19	-0.11	0.22	0.07	0.08	0.41	0.26	0.33	0.18	-0.15
-0.35	-0.55	-0.51	-0.18	-0.34	-0.4	-0.36	-0.03	-0.19	-0.18	0.15	-0.01	0.04	-0.12	-0.46
0.06	-0.11	0	0.34	0.2	0.03	0.14	0.48	0.34	0.34	0.67	0.53	0.63	0.49	0.16
Total														
Number of Types														
0.68	0.63	0.96	0.72	1	1	0.31	0.13	0.74	0.27	0.12	0.7	0.99	1	0.97
2.78	3.04	-1.83	-3.68	-1.05	0.27	-4.6	-6.46	-3.83	-4.87	-6.73	-4.09	-1.86	0.78	2.63
-2.66	-2.6	-8.26	-11.17	-9.07	-5.32	-10.99	-13.91	-11.81	-11.44	-14.33	-12.22	-10.06	-7.92	-6.86
8.21	8.69	4.61	3.8	6.97	5.86	1.79	0.99	4.15	1.7	0.87	4.03	6.35	9.47	12.13
Cyclomatic Complexity (CC)														
Maximum														
1	0.67	0.35	0.7	0.41	0.44	0.2	0.53	0.56	0.98	1	0.05	1	0.02	0.07
0.54	-2.35	-3.59	-3.03	4.24	-2.89	-4.13	-3.57	3.7	-1.23	-0.67	6.59	0.56	7.83	7.27
-3.83	-6.9	-8.77	-9.06	-2.22	-7.4	-9.27	-9.57	-2.73	-6.52	-6.8	0.05	-6.05	0.83	-0.38
4.92	2.19	1.6	3	10.7	1.61	1.02	2.43	10.12	4.05	5.45	13.13	7.17	14.83	14.91
Standard Deviation														
0.99	0.39	0.39	0.61	0.71	0.78	0.74	0.88	0.41	1	1	0.07	1	0.07	0.15
-0.13	-0.41	-0.46	-0.44	0.42	-0.27	-0.33	-0.31	0.56	-0.06	-0.03	0.83	0.02	0.89	0.86
-0.72	-1.01	-1.15	-1.24	-0.43	-0.87	-1.01	-1.1	-0.29	-0.76	-0.85	-0.04	-0.86	-0.04	-0.15
0.45	0.2	0.23	0.36	1.28	0.33	0.36	0.49	1.41	0.65	0.78	1.7	0.9	1.82	1.88
Number Of Fields (NOF)														
Alves Medium Risk														
0.32	0.97	0.97	0.39	1	0.84	0.92	0.01	0.83	1	0.14	1	0.18	1	0.56
-0.03	-0.01	-0.01	0.03	0	0.02	0.01	0.06	0.02	0	0.04	0.01	0.04	0.01	-0.04
-0.06	-0.05	-0.05	-0.02	-0.06	-0.02	-0.03	0.01	-0.03	-0.04	-0.01	-0.05	-0.01	-0.05	-0.1
0.01	0.03	0.03	0.08	0.05	0.05	0.06	0.11	0.08	0.04	0.09	0.06	0.1	0.07	0.03

B-A	C-A	D-A	E-A	F-A	C-B	D-B	E-B	F-B	D-C	E-C	F-C	E-D	F-D	F-E
pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl
dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif
lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob
upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb
Number Of Private Fields (NOPF)														
Mean														
1	1	0.05	1	1	1	0.02	1	0.99	0.05	1	1	0.15	0.46	1
-0.05	0	0.59	-0.04	0.1	0.05	0.64	0.01	0.15	0.59	-0.03	0.1	-0.63	-0.49	0.13
-0.54	-0.52	0.01	-0.71	-0.63	-0.46	0.06	-0.66	-0.58	0	-0.72	-0.64	-1.37	-1.28	-0.73
0.44	0.51	1.17	0.64	0.82	0.55	1.22	0.69	0.87	1.19	0.66	0.84	0.12	0.29	0.99
Alves Very High Risk														
0.98	1	0.05	1	1	0.99	0.01	0.99	0.93	0.05	1	0.99	0.25	0.54	1
-0.02	0	0.08	0	0.02	0.02	0.1	0.02	0.03	0.08	0	0.02	-0.08	-0.06	0.01
-0.09	-0.07	0	-0.09	-0.08	-0.05	0.02	-0.07	-0.07	0	-0.09	-0.08	-0.18	-0.17	-0.1
0.05	0.07	0.16	0.1	0.12	0.09	0.18	0.11	0.13	0.16	0.1	0.12	0.02	0.05	0.13
Number Of Private Methods (NOPM)														
Alves Low Risk														
0.97	1	0.79	0.96	0	0.92	0.99	1	0.02	0.67	0.91	0	1	0.14	0.14
-0.01	0	-0.02	-0.02	-0.08	0.02	-0.01	-0.01	-0.07	-0.03	-0.02	-0.09	0.01	-0.06	-0.07
-0.06	-0.04	-0.08	-0.08	-0.15	-0.03	-0.06	-0.07	-0.14	-0.08	-0.08	-0.15	-0.06	-0.13	-0.14
0.03	0.05	0.03	0.04	-0.02	0.06	0.04	0.05	-0.01	0.03	0.04	-0.02	0.07	0.01	0.01
Type Lines of Code (TLOC)														
Variance														
1	0.95	0.63	0.99	0.14	1	0.42	1	0.07	0.21	1	0.03	0.47	0.87	0.1
-231.83	-579.32	1180.2	-558.33	2339	-347.48	1412.0	-326.5	2570.8	1759.5	20.98	2918.3	-1738.5	1158.7	2897.34
-2073.6	-2491.6	-1001.0	-3096.0	-377.95	-2241.9	-753.57	-2850.7	-133.61	-466.3	-2555.1	165.42	-4520.2	-1787.3	-321.59
1610.0	1332.9	3361.4	1979.3	5055.9	1546.9	3577.7	2197.7	5275.2	3985.4	2597.1	5671.2	1043.0	4104.8	6116.27
Standard Deviation														
1	0.77	0.81	0.99	0.36	0.87	0.69	1	0.27	0.17	1	0.05	0.66	0.95	0.28
-1.07	-6.31	6.77	-3.58	13.28	-5.24	7.84	-2.5	14.35	13.08	2.74	19.59	-10.34	6.51	16.85
-14.2	-19.94	-8.77	-21.66	-6.08	-18.74	-7.59	-20.49	-4.92	-2.78	-15.62	-0.03	-30.17	-14.48	-6.08
12.05	7.31	22.31	14.51	32.64	8.26	23.28	15.49	33.62	28.94	21.09	39.21	9.48	27.5	39.79
Weighted Methods per Class (WMC)														
Variance														
1	1	1	0.92	0.04	0.95	1	0.74	0.09	0.99	0.98	0.02	0.89	0.1	0.01
9.47	-7.7	4.74	-26.27	83.89	-17.16	-4.73	-35.73	74.42	12.44	-18.57	91.58	-31.01	79.15	110.15
-45.29	-64.55	-60.11	-101.71	3.11	-73.49	-69.11	-110.78	-5.98	-53.74	-95.16	9.74	-113.7	-8.44	14.45
64.22	49.15	69.59	49.18	164.66	39.16	59.66	39.32	154.82	78.61	58.02	173.43	51.69	166.73	205.85
Number of Children (NC)														
Maximum														
0.36	0.17	0.93	1	1	1	0.09	0.71	0.77	0.04	0.49	0.58	0.98	0.98	1
0.87	1.09	-0.49	0	0.01	0.22	-1.36	-0.87	-0.85	-1.58	-1.09	-1.08	0.5	0.51	0.01
-0.4	-0.23	-2	-1.75	-1.86	-1.08	-2.85	-2.6	-2.72	-3.12	-2.86	-2.97	-1.42	-1.52	-2.2
2.13	2.41	1.01	1.75	1.88	1.53	0.13	0.87	1.01	-0.05	0.68	0.82	2.41	2.53	2.23

B-A	C-A	D-A	E-A	F-A	C-B	D-B	E-B	F-B	D-C	E-C	F-C	E-D	F-D	F-E
pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl
dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif
lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob
upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb
Lack of Cohesion Of Methods (LCOM)														
Percentile Medium Risk														
1	1	0.99	0.48	0.01	1	1	0.5	0.02	1	0.73	0.05	0.85	0.1	0.74
0	0.01	0.01	0.04	0.08	0.01	0.01	0.04	0.08	0	0.03	0.07	0.03	0.07	0.04
-0.05	-0.04	-0.05	-0.03	0.01	-0.04	-0.05	-0.03	0.01	-0.06	-0.03	0	-0.04	-0.01	-0.04
0.05	0.06	0.07	0.11	0.15	0.06	0.07	0.11	0.15	0.06	0.1	0.15	0.1	0.15	0.12
Alves Low Risk														
0.69	1	1	1	0.21	0.81	0.93	0.67	0.01	1	0.99	0.17	0.99	0.19	0.61
0.04	0.01	0.01	-0.02	-0.1	-0.04	-0.03	-0.06	-0.15	0.01	-0.02	-0.11	-0.03	-0.11	-0.09
-0.04	-0.09	-0.09	-0.14	-0.23	-0.13	-0.14	-0.18	-0.28	-0.1	-0.15	-0.24	-0.16	-0.25	-0.24
0.13	0.1	0.12	0.1	0.03	0.05	0.07	0.06	-0.02	0.11	0.1	0.02	0.1	0.03	0.07
Fan-in (FANIN)														
Alves Very High Risk														
0.5	1	1	0.03	0.49	0.79	0.41	0.43	0.99	0.98	0.07	0.69	0.02	0.39	0.93
0.01	0	0	0.02	0.01	-0.01	-0.01	0.01	0	-0.01	0.02	0.01	0.03	0.02	-0.01
-0.01	-0.01	-0.02	0	-0.01	-0.02	-0.03	-0.01	-0.02	-0.02	0	-0.01	0	-0.01	-0.04
0.03	0.02	0.02	0.05	0.04	0.01	0.01	0.04	0.03	0.01	0.05	0.04	0.05	0.04	0.02

Table 4.6: Tukey differences for TA teams Categorisation B on all data

B-A	**C-A**	**D-A**	**E-A**	**F-A**	**C-B**	**D-B**	**E-B**	**F-B**	**D-C**	**E-C**	**F-C**	**E-D**	**F-D**	**F-E**
pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl
dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif
lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob
upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb
Resultant Variable														
OO Standard														
0.28	0	0.73	0.08	0.8	0.37	0.02	0.92	1	0	0.99	0.41	0.01	0.2	0.86
-0.15	-0.3	0.12	-0.24	-0.12	-0.14	0.28	-0.09	0.03	0.42	0.06	0.18	-0.36	-0.24	0.12
-0.36	-0.51	-0.13	-0.5	-0.39	-0.35	0.03	-0.34	-0.23	0.16	-0.21	-0.09	-0.66	-0.54	-0.18
0.05	-0.08	0.38	0.02	0.15	0.07	0.52	0.17	0.3	0.68	0.32	0.45	-0.07	0.06	0.43
Coding Standard														
0.33	0	0.05	0.95	0.97	0.12	0.81	0.12	0.97	0.95	0	0.07	0.02	0.51	0.72
-0.15	-0.33	-0.25	0.08	-0.07	-0.19	-0.11	0.22	0.07	0.08	0.41	0.26	0.33	0.18	-0.15
-0.35	-0.55	-0.51	-0.18	-0.34	-0.4	-0.36	-0.03	-0.19	-0.18	0.15	-0.01	0.04	-0.12	-0.46
0.06	-0.11	0	0.34	0.2	0.03	0.14	0.48	0.34	0.34	0.67	0.53	0.63	0.49	0.16
Total														
Number of Types														
0.68	0.63	0.68	1	0.9	1	0.11	0.71	0.27	0.1	0.66	0.24	0.87	1	0.98
2.78	3.04	-3.85	-0.56	-2.66	0.27	-6.63	-3.34	-5.43	-6.89	-3.6	-5.7	3.29	1.19	-2.1
-2.65	-2.59	-11.33	-7.32	-9.99	-5.31	-14.06	-10.05	-12.72	-14.48	-10.49	-13.15	-5.17	-7.73	-10.43
8.2	8.68	3.63	6.2	4.67	5.85	0.81	3.38	1.86	0.7	3.28	1.74	11.75	10.11	6.23
Method Lines of Code (MLOC)														
Percentile Low Risk														
0.82	0.7	0.98	1	0.06	1	0.61	0.95	0.41	0.5	0.89	0.56	0.98	0.05	0.18
0.02	0.03	-0.02	0	0.07	0	-0.04	-0.02	0.05	-0.04	-0.02	0.04	0.02	0.09	0.07
-0.03	-0.03	-0.09	-0.06	0	-0.05	-0.11	-0.08	-0.02	-0.12	-0.09	-0.03	-0.06	0	-0.01
0.07	0.08	0.06	0.07	0.14	0.06	0.03	0.05	0.12	0.03	0.04	0.11	0.1	0.17	0.15

B-A	**C-A**	**D-A**	**E-A**	**F-A**	**C-B**	**D-B**	**E-B**	**F-B**	**D-C**	**E-C**	**F-C**	**E-D**	**F-D**	**F-E**
pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl
dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif
lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob
upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb
Number Of Private Fields (NOPF)														
Median														
1	1	0.01	1	1	1	0.01	1	1	0.02	1	1	0.05	0.04	1
0.02	0.03	0.61	0.04	0	0.01	0.59	0.02	-0.02	0.58	0.01	-0.03	-0.57	-0.61	-0.04
-0.34	-0.34	0.11	-0.41	-0.49	-0.36	0.09	-0.43	-0.51	0.07	-0.45	-0.53	-1.14	-1.21	-0.6
0.38	0.41	1.11	0.49	0.49	0.39	1.09	0.47	0.47	1.09	0.47	0.46	0	-0.01	0.52
Mean														
1	1	0.01	1	1	1	0	1	1	0.01	1	1	0.03	0.05	1
-0.05	0	0.84	0.05	0.05	0.05	0.89	0.1	0.1	0.85	0.05	0.05	-0.79	-0.79	0
-0.54	-0.51	0.17	-0.56	-0.61	-0.46	0.23	-0.5	-0.55	0.17	-0.56	-0.61	-1.55	-1.59	-0.75
0.44	0.5	1.51	0.66	0.71	0.55	1.56	0.7	0.75	1.53	0.67	0.72	-0.04	0.01	0.75
Variance														
1	1	0.03	1	0.97	1	0.02	1	0.94	0.02	1	0.95	0.07	0.38	0.98
-0.32	-0.27	7.27	-0.1	1.79	0.05	7.6	0.22	2.11	7.55	0.17	2.06	-7.38	-5.48	1.89
-5.28	-5.42	0.44	-6.28	-4.91	-5.05	0.8	-5.92	-4.55	0.61	-6.12	-4.74	-15.1	-13.63	-5.71
4.63	4.87	14.1	6.07	8.49	5.15	14.39	6.35	8.77	14.48	6.46	8.86	0.35	2.67	9.5
Standard Deviation														
0.96	1	0.07	1	0.98	1	0.01	0.99	0.75	0.04	1	0.92	0.12	0.53	0.97
-0.18	-0.08	0.78	-0.04	0.2	0.1	0.95	0.14	0.38	0.86	0.04	0.28	-0.82	-0.58	0.24
-0.77	-0.69	-0.04	-0.78	-0.6	-0.51	0.14	-0.6	-0.42	0.03	-0.71	-0.54	-1.74	-1.55	-0.67
0.42	0.54	1.6	0.7	1	0.71	1.77	0.87	1.17	1.69	0.79	1.09	0.11	0.4	1.15
Alves Very High Risk														
0.98	1	0.01	0.99	1	0.99	0	0.82	0.99	0.01	0.98	1	0.11	0.05	1
-0.02	0	0.11	0.02	0	0.02	0.13	0.04	0.02	0.12	0.02	0	-0.09	-0.11	-0.02
-0.08	-0.07	0.02	-0.06	-0.09	-0.05	0.04	-0.05	-0.07	0.02	-0.06	-0.09	-0.2	-0.22	-0.12
0.05	0.07	0.21	0.1	0.09	0.08	0.22	0.12	0.11	0.21	0.11	0.1	0.01	0	0.09
Number Of Private Methods (NOPM)														
Alves Low Risk														
0.97	1	0.19	0.99	0.1	0.92	0.48	1	0.31	0.14	0.96	0.07	0.64	1	0.48
-0.01	0	-0.05	-0.01	-0.05	0.02	-0.04	0	-0.04	-0.05	-0.02	-0.06	0.04	-0.01	-0.04
-0.06	-0.04	-0.11	-0.07	-0.11	-0.03	-0.1	-0.06	-0.1	-0.12	-0.07	-0.12	-0.03	-0.08	-0.11
0.03	0.05	0.01	0.04	0.01	0.06	0.02	0.05	0.02	0.01	0.04	0	0.11	0.07	0.03

B-A	C-A	D-A	E-A	F-A	C-B	D-B	E-B	F-B	D-C	E-C	F-C	E-D	F-D	F-E
pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl
dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif
lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob
upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb
Type Lines of Code (TLOC)														
Variance														
1	0.95	0.11	0.99	0.49	0.99	0.05	1	0.32	0.02	1	0.16	0.07	0.98	0.32
-231.83	-579.32	2261.9	-482.6	1523.6	-347.48	2493.8	-250.77	1755.4	2841.3	96.71	2102.9	-2744.5	-738.38	2006.21
-2066.8	-2484.5	-266.28	-2768.0	-955.52	-2234.9	-21.09	-2521.4	-710.07	274.73	-2231.0	-415.25	-5604.8	-3755.6	-810.69
1603.1	1325.8	4790.2	1802.8	4002.7	1539.9	5008.7	2019.9	4220.9	5407.8	2424.4	4621.1	115.65	2278.8	4823.11
Standard Deviation														
1	0.76	0.25	1	0.84	0.87	0.17	1	0.74	0.02	0.99	0.25	0.21	0.96	0.72
-1.07	-6.31	13.67	-2.52	7.31	-5.24	14.74	-1.45	8.38	19.98	3.79	13.62	-16.2	-6.36	9.83
-14.17	-19.91	-4.38	-18.84	-10.39	-18.71	-3.21	-17.66	-9.22	1.66	-12.83	-4.36	-36.61	-27.9	-10.28
12.03	7.29	31.72	13.79	25.01	8.24	32.7	14.76	25.99	38.31	20.41	31.6	4.22	15.18	29.94
Alves Very High Risk														
1	1	0.25	0.92	0.29	1	0.19	0.95	0.22	0.23	0.95	0.27	0.07	1	0.09
0	0	0.02	-0.01	0.02	0	0.02	-0.01	0.02	0.02	-0.01	0.02	-0.03	0	0.02
-0.02	-0.02	-0.01	-0.03	-0.01	-0.02	0	-0.03	-0.01	-0.01	-0.03	-0.01	-0.05	-0.03	0
0.02	0.02	0.04	0.01	0.04	0.02	0.04	0.01	0.04	0.04	0.01	0.04	0	0.03	0.05
Number of Children (NC)														
Maximum														
0.37	0.17	1	0.99	1	1	0.71	0.23	0.38	0.49	0.11	0.21	1	1	1
0.87	1.09	0	-0.35	-0.28	0.22	-0.87	-1.21	-1.15	-1.09	-1.44	-1.37	-0.35	-0.28	0.06
-0.4	-0.23	-1.75	-1.93	-2	-1.08	-2.61	-2.78	-2.85	-2.86	-3.05	-3.11	-2.33	-2.37	-1.88
2.14	2.41	1.75	1.23	1.43	1.53	0.87	0.36	0.56	0.69	0.17	0.37	1.63	1.8	2.01

Table 4.7: Tukey differences for TA teams Categorisation A on complete data

B-A	**C-A**	**D-A**	**E-A**	**F-A**	**C-B**	**D-B**	**E-B**	**F-B**	**D-C**	**E-C**	**F-C**	**E-D**	**F-D**	**F-E**
pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl
dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif
lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob
upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb
Resultant Variable														
OO Standard														
0.19	0.09	1	0.17	1	1	0.57	0.82	0.81	0.39	0.94	0.62	0.27	1	0.42
-0.19	-0.23	0.02	-0.38	-0.04	-0.04	0.21	-0.2	0.14	0.25	-0.15	0.19	-0.41	-0.07	0.34
-0.42	-0.48	-0.35	-0.85	-0.38	-0.29	-0.16	-0.66	-0.19	-0.13	-0.63	-0.16	-0.95	-0.51	-0.19
0.05	0.02	0.39	0.08	0.29	0.21	0.58	0.27	0.48	0.63	0.32	0.53	0.14	0.38	0.87
Coding Standard														
0.28	0.01	0.16	1	1	0.76	0.89	0.69	0.8	1	0.27	0.25	0.37	0.43	1
-0.17	-0.28	-0.3	0.06	-0.02	-0.11	-0.13	0.23	0.14	-0.02	0.35	0.26	0.37	0.28	-0.09
-0.4	-0.53	-0.67	-0.4	-0.36	-0.36	-0.5	-0.23	-0.19	-0.4	-0.12	-0.08	-0.18	-0.16	-0.61
0.06	-0.04	0.06	0.53	0.31	0.13	0.23	0.7	0.48	0.36	0.82	0.6	0.91	0.71	0.43
Total														
Number of Types														
0.75	0.12	0.89	0.93	1	0.81	0.3	1	0.94	0.04	1	0.55	0.66	0.99	0.95
2.88	5.69	-3.07	4.44	-0.43	2.81	-5.96	1.56	-3.31	-8.77	-1.25	-6.12	7.51	2.65	-4.87
-3.24	-0.83	-11.31	-8.82	-10.72	-3.71	-14.19	-11.7	-13.61	-17.31	-14.71	-16.66	-6.85	-9.03	-20.5
9	12.21	5.16	17.7	9.87	9.33	2.28	14.82	6.99	-0.23	12.2	4.42	21.87	14.33	10.77
Parameter Count (PC)														
Median														
0.91	1	0.73	0.97	0.08	0.83	0.27	1	0.29	0.86	0.94	0.06	0.64	0.01	0.88
-0.1	0.03	0.18	-0.16	-0.43	0.12	0.27	-0.07	-0.33	0.15	-0.19	-0.45	-0.34	-0.6	-0.27
-0.37	-0.26	-0.19	-0.75	-0.89	-0.17	-0.09	-0.65	-0.79	-0.23	-0.78	-0.92	-0.97	-1.12	-0.96
0.18	0.32	0.54	0.43	0.03	0.41	0.64	0.52	0.12	0.53	0.41	0.01	0.3	-0.09	0.43
Number Of Fields (NOF)														
Percentile Medium Risk														
0.99	0.99	0.6	0.41	0.3	1	0.3	0.6	0.54	0.37	0.6	0.54	0.09	0.04	1
-0.02	-0.02	0.06	-0.11	-0.09	0	0.08	-0.09	-0.08	0.07	-0.09	-0.08	-0.17	-0.15	0.02
-0.09	-0.1	-0.05	-0.28	-0.22	-0.08	-0.03	-0.26	-0.21	-0.03	-0.26	-0.21	-0.35	-0.3	-0.18
0.06	0.07	0.16	0.06	0.04	0.08	0.18	0.08	0.05	0.18	0.08	0.06	0.01	0	0.21

B-A	**C-A**	**D-A**	**E-A**	**F-A**	**C-B**	**D-B**	**E-B**	**F-B**	**D-C**	**E-C**	**F-C**	**E-D**	**F-D**	**F-E**
pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl
dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif
lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob
upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb
Number Of Private Fields (NOPF)														
Maximum														
0.84	1	0.58	0.97	0.24	0.94	0.13	1	0.05	0.52	0.99	0.21	0.57	0.97	0.29
-1.1	-0.2	2.01	-1.53	3.4	0.89	3.1	-0.43	4.5	2.21	-1.33	3.61	-3.54	1.4	4.93
-3.74	-3.02	-1.55	-7.26	-1.05	-1.93	-0.46	-6.17	0.05	-1.48	-7.15	-0.95	-9.75	-3.66	-1.83
1.55	2.62	5.57	4.21	7.86	3.71	6.67	5.3	8.95	5.9	4.49	8.17	2.67	6.45	11.7
Median														
1	1	0.02	1	1	1	0.03	1	1	0.04	1	1	0.46	0.23	1
0.02	0.02	0.69	0	0	-0.01	0.66	-0.02	-0.02	0.67	-0.02	-0.02	-0.69	-0.69	0
-0.44	-0.48	0.06	-1.01	-0.78	-0.5	0.04	-1.03	-0.81	0.02	-1.04	-0.82	-1.78	-1.58	-1.19
0.49	0.51	1.31	1.01	0.78	0.49	1.29	0.98	0.76	1.32	1.01	0.79	0.4	0.2	1.19
Mean														
1	1	0.02	1	1	1	0.01	1	0.98	0.03	1	1	0.31	0.46	0.99
-0.07	0	0.91	-0.12	0.17	0.07	0.98	-0.05	0.24	0.91	-0.12	0.17	-1.03	-0.74	0.29
-0.68	-0.65	0.08	-1.45	-0.86	-0.58	0.15	-1.38	-0.79	0.05	-1.47	-0.89	-2.47	-1.91	-1.28
0.54	0.66	1.74	1.21	1.21	0.73	1.81	1.28	1.28	1.76	1.23	1.23	0.41	0.43	1.86
Standard Deviation														
0.91	1	0.16	0.99	0.59	0.98	0.02	1	0.23	0.12	1	0.5	0.35	1	0.58
-0.25	-0.07	0.8	-0.35	0.67	0.18	1.05	-0.1	0.93	0.87	-0.28	0.74	-1.15	-0.13	1.03
-0.96	-0.83	-0.16	-1.9	-0.52	-0.57	0.09	-1.64	-0.27	-0.12	-1.85	-0.48	-2.82	-1.48	-0.79
0.46	0.69	1.76	1.19	1.87	0.94	2.01	1.44	2.12	1.86	1.28	1.97	0.52	1.23	2.84
Alves Medium Risk														
0.74	0.9	0.17	1	1	1	0.01	0.95	0.99	0.03	0.98	1	0.86	0.41	1
0.02	0.01	-0.04	-0.01	0.01	0	-0.06	-0.02	-0.01	-0.05	-0.02	-0.01	0.03	0.05	0.01
-0.02	-0.02	-0.09	-0.08	-0.06	-0.04	-0.11	-0.1	-0.07	-0.1	-0.1	-0.07	-0.05	-0.02	-0.08
0.05	0.05	0.01	0.07	0.07	0.04	-0.01	0.05	0.05	0	0.06	0.05	0.12	0.11	0.1
Alves Very High Risk														
1	1	0	0.99	1	1	0	1	0.99	0	0.99	1	0.08	0.15	0.98
-0.01	0	0.13	-0.03	0.01	0.01	0.14	-0.02	0.02	0.13	-0.03	0.01	-0.16	-0.12	0.04
-0.09	-0.08	0.03	-0.19	-0.11	-0.07	0.04	-0.18	-0.1	0.03	-0.2	-0.12	-0.34	-0.26	-0.15
0.06	0.08	0.23	0.13	0.14	0.09	0.24	0.14	0.15	0.24	0.13	0.14	0.01	0.02	0.24
Number Of Methods (NOM)														
Percentile High Risk														
0.56	1	0.36	0.96	0.98	0.34	0.02	1	1	0.63	0.89	0.92	0.4	0.34	1
-0.03	0.01	0.05	-0.04	-0.02	0.04	0.09	0	0.01	0.04	-0.05	-0.03	-0.09	-0.08	0.01
-0.09	-0.05	-0.03	-0.17	-0.12	-0.02	0.01	-0.13	-0.09	-0.04	-0.18	-0.14	-0.23	-0.19	-0.14
0.03	0.07	0.13	0.09	0.08	0.11	0.17	0.13	0.11	0.13	0.08	0.07	0.05	0.03	0.17
Alves Very High Risk														
0.5	0.83	0.95	0.97	0.01	1	0.25	0.59	0.19	0.48	0.74	0.12	1	0.01	0.05
0.01	0.01	-0.01	-0.01	0.04	0	-0.02	-0.02	0.02	-0.02	-0.02	0.03	0	0.04	0.05
-0.01	-0.01	-0.03	-0.05	0.01	-0.02	-0.04	-0.06	-0.01	-0.04	-0.06	0	-0.05	0.01	0
0.03	0.03	0.02	0.03	0.07	0.02	0.01	0.02	0.05	0.01	0.02	0.06	0.04	0.08	0.09

B-A	**C-A**	**D-A**	**E-A**	**F-A**	**C-B**	**D-B**	**E-B**	**F-B**	**D-C**	**E-C**	**F-C**	**E-D**	**F-D**	**F-E**
pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl
dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif
lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob
upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb
Number Of Private Methods (NOPM)														
Minimum														
0.91	0.97	0.09	0.79	0.28	1	0.01	0.96	0.66	0.02	0.95	0.61	0.08	0	1
-0.12	-0.1	0.41	-0.32	-0.41	0.02	0.53	-0.2	-0.29	0.51	-0.22	-0.31	-0.74	-0.83	-0.09
-0.45	-0.45	-0.03	-1.05	-0.97	-0.34	0.08	-0.93	-0.85	0.05	-0.96	-0.89	-1.52	-1.46	-0.94
0.21	0.26	0.86	0.4	0.15	0.37	0.98	0.52	0.27	0.98	0.51	0.26	0.04	-0.19	0.76
Type Lines of Code (TLOC)														
Percentile Medium Risk														
0.58	0.95	0.38	1	1	0.15	0.02	0.95	1	0.84	1	0.9	0.91	0.46	1
-0.04	0.02	0.06	0.01	-0.02	0.06	0.09	0.04	0.02	0.04	-0.02	-0.04	-0.05	-0.08	-0.02
-0.1	-0.05	-0.03	-0.13	-0.13	-0.01	0.01	-0.1	-0.09	-0.05	-0.16	-0.15	-0.2	-0.2	-0.19
0.03	0.09	0.14	0.14	0.09	0.13	0.18	0.18	0.12	0.13	0.12	0.07	0.1	0.05	0.14
Weighted Methods per Class (WMC)														
Variance														
1	0.99	1	0.89	0.08	0.9	0.98	0.77	0.16	1	0.97	0.04	0.97	0.09	0.08
11.19	-13.26	-10	-50.71	99.15	-24.45	-21.2	-61.9	87.96	3.25	-37.45	112.41	-40.7	109.16	149.86
-51.28	-79.86	-94.12	-186.16	-6.01	-91.05	-105.31	-197.35	-17.21	-83.97	-174.85	4.74	-187.4	-10.14	-9.84
73.67	53.35	74.11	84.74	204.32	42.15	62.92	73.55	193.13	90.47	99.95	220.08	105.99	228.46	309.56
Number of Children (NC)														
Maximum														
0.64	0.06	0.86	0.97	0.99	0.75	0.21	1	1	0.02	0.99	0.88	0.74	0.74	1
0.81	1.57	-0.81	0.89	0.58	0.76	-1.62	0.08	-0.23	-2.38	-0.68	-0.99	1.7	1.39	-0.31
-0.71	-0.05	-2.85	-2.4	-1.98	-0.86	-3.66	-3.21	-2.79	-4.5	-4.02	-3.61	-1.87	-1.51	-4.19
2.33	3.19	1.24	4.18	3.14	2.38	0.43	3.37	2.33	-0.26	2.66	1.63	5.27	4.29	3.57
Alves Medium Risk														
0.67	1	0.13	0.98	0.79	0.53	0.73	0.71	1	0.09	0.99	0.7	0.28	0.99	0.68
0.02	0	0.05	-0.02	0.03	-0.03	0.03	-0.05	0.01	0.06	-0.02	0.04	-0.07	-0.02	0.06
-0.02	-0.05	-0.01	-0.12	-0.04	-0.07	-0.03	-0.14	-0.06	0	-0.11	-0.04	-0.18	-0.1	-0.05
0.07	0.04	0.11	0.07	0.11	0.02	0.09	0.05	0.08	0.12	0.08	0.11	0.03	0.06	0.17
Depth of Inheritance Tree (DIT)														
Alves Low Risk														
0.7	0.24	0.86	0.21	0.99	0.96	0.23	0.6	1	0.06	0.85	0.96	0.07	0.79	0.62
-0.06	-0.09	0.06	-0.2	-0.04	-0.04	0.12	-0.14	0.02	0.16	-0.1	0.06	-0.26	-0.1	0.16
-0.18	-0.22	-0.09	-0.45	-0.23	-0.16	-0.04	-0.39	-0.17	-0.01	-0.36	-0.14	-0.54	-0.32	-0.14
0.06	0.03	0.22	0.05	0.16	0.09	0.28	0.11	0.22	0.32	0.15	0.26	0.01	0.12	0.46
Fan-in (FANIN)														
Alves Very High Risk														
0.64	0.98	0.26	0.11	0.45	0.98	0.01	0.43	0.94	0.1	0.25	0.75	0.01	0.02	0.93
0.01	0	-0.02	0.03	0.02	0	-0.03	0.02	0.01	-0.02	0.03	0.01	0.05	0.03	-0.01
-0.01	-0.01	-0.04	0	-0.01	-0.02	-0.05	-0.01	-0.02	-0.04	-0.01	-0.01	0.01	0	-0.06
0.03	0.02	0.01	0.07	0.05	0.01	0	0.06	0.04	0	0.06	0.04	0.09	0.07	0.03

Table 4.8: Tukey differences for TA teams Categorisation B on complete data

B-A	C-A	D-A	E-A	F-A	C-B	D-B	E-B	F-B	D-C	E-C	F-C	E-D	F-D	F-E
pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl
dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif
lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob
upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb
Resultant Variable														
OO Standard														
0.19	0.09	1	0.17	1	1	0.57	0.82	0.81	0.39	0.94	0.62	0.27	1	0.42
-0.19	-0.23	0.02	-0.38	-0.04	-0.04	0.21	-0.2	0.14	0.25	-0.15	0.19	-0.41	-0.07	0.34
-0.42	-0.48	-0.35	-0.85	-0.38	-0.29	-0.16	-0.66	-0.19	-0.13	-0.63	-0.16	-0.95	-0.51	-0.19
0.05	0.02	0.39	0.08	0.29	0.21	0.58	0.27	0.48	0.63	0.32	0.53	0.14	0.38	0.87
Coding Standard														
0.28	0.01	0.16	1	1	0.76	0.89	0.69	0.8	1	0.27	0.25	0.37	0.43	1
-0.17	-0.28	-0.3	0.06	-0.02	-0.11	-0.13	0.23	0.14	-0.02	0.35	0.26	0.37	0.28	-0.09
-0.4	-0.53	-0.67	-0.4	-0.36	-0.36	-0.5	-0.23	-0.19	-0.4	-0.12	-0.08	-0.18	-0.16	-0.61
0.06	-0.04	0.06	0.53	0.31	0.13	0.23	0.7	0.48	0.36	0.82	0.6	0.91	0.71	0.43
Total														
Number of Methods														
0.73	0.4	0.19	0.47	0.99	0.99	0.03	0.94	0.6	0.01	1	0.36	0.02	0.61	0.35
14.93	21.81	-46.33	31.04	-9.92	6.89	-61.26	16.11	-24.84	-68.15	9.22	-31.73	77.37	36.42	-40.95
-15.96	-11.11	-104.12	-18.77	-54.84	-26.04	-119.05	-33.7	-69.77	-127.05	-41.87	-78.08	7.61	-29.94	-100.5
45.82	54.74	11.46	80.85	35.01	39.81	-3.47	65.92	20.08	-9.24	60.32	14.62	147.13	102.78	18.59
Number of Types														
0.72	0.1	0.22	0.28	0.9	0.79	0.04	0.81	0.33	0	1	0.05	0.01	0.81	0.11
2.88	5.69	-8.62	7.04	-3.15	2.81	-11.5	4.16	-6.03	-14.31	1.35	-8.84	15.66	5.47	-10.18
-3.03	-0.6	-19.67	-2.49	-11.74	-3.49	-22.55	-5.37	-14.62	-25.58	-8.43	-17.7	2.32	-7.22	-21.57
8.79	11.99	2.43	16.56	5.45	9.11	-0.45	13.68	2.56	-3.05	11.12	0.03	29	18.16	1.2
Method Lines of Code (MLOC)														
Mean														
0.53	0.11	0.6	0.56	0.29	0.93	0.13	0.99	0.94	0.04	1	1	0.13	0.06	1
-0.71	-1.14	1.24	-1.12	-1.28	-0.43	1.95	-0.41	-0.57	2.39	0.03	-0.14	-2.36	-2.52	-0.16
-1.91	-2.43	-1.01	-3.06	-3.03	-1.72	-0.3	-2.35	-2.32	0.09	-1.97	-1.95	-5.08	-5.11	-2.49
0.5	0.14	3.5	0.83	0.47	0.85	4.21	1.54	1.18	4.68	2.02	1.67	0.36	0.06	2.16
Percentile Low Risk														
0.86	0.42	0.87	1	0.12	0.97	0.51	0.92	0.5	0.27	0.67	0.87	0.98	0.08	0.27
0.02	0.04	-0.04	-0.01	0.07	0.02	-0.06	-0.03	0.05	-0.08	-0.05	0.03	0.03	0.11	0.08
-0.03	-0.02	-0.14	-0.1	-0.01	-0.04	-0.17	-0.12	-0.03	-0.18	-0.14	-0.05	-0.09	-0.01	-0.03
0.08	0.1	0.06	0.08	0.15	0.08	0.04	0.06	0.13	0.03	0.04	0.12	0.16	0.23	0.19
Percentile Medium Risk														
0.96	0.77	0.81	0.89	0.14	1	0.53	0.6	0.42	0.36	0.41	0.7	1	0.07	0.07
-0.01	-0.02	0.03	0.02	-0.04	-0.01	0.04	0.03	-0.03	0.04	0.04	-0.03	-0.01	-0.07	-0.06
-0.04	-0.05	-0.04	-0.03	-0.09	-0.04	-0.03	-0.02	-0.08	-0.02	-0.02	-0.08	-0.08	-0.14	-0.13
0.02	0.02	0.09	0.07	0.01	0.03	0.1	0.08	0.02	0.11	0.09	0.03	0.07	0	0

B-A	**C-A**	**D-A**	**E-A**	**F-A**	**C-B**	**D-B**	**E-B**	**F-B**	**D-C**	**E-C**	**F-C**	**E-D**	**F-D**	**F-E**
pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl
dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif
lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob
upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb
Number Of Private Fields (NOPF)														
Minimum														
1	1	0	1	1	1	0	1	1	0	1	1	0	0	1
0	0	0.14	0	0	0	0.14	0	0	0.14	0	0	-0.14	-0.14	0
-0.05	-0.05	0.05	-0.08	-0.07	-0.05	0.05	-0.08	-0.07	0.05	-0.08	-0.07	-0.25	-0.25	-0.09
0.05	0.05	0.23	0.08	0.07	0.05	0.23	0.08	0.07	0.24	0.08	0.07	-0.03	-0.04	0.09
Maximum														
0.84	1	0.12	1	0.74	0.94	0.02	0.97	0.24	0.1	1	0.67	0.31	0.8	0.92
-1.1	-0.2	4.36	0.07	1.84	0.89	5.45	1.17	2.94	4.56	0.27	2.04	-4.29	-2.52	1.77
-3.74	-3.02	-0.6	-4.2	-2.01	-1.93	0.5	-3.1	-0.92	-0.49	-4.11	-1.93	-10.27	-8.21	-3.33
1.55	2.62	9.31	4.34	5.69	3.72	10.41	5.44	6.79	9.61	4.65	6.02	1.69	3.17	6.87
Median														
1	1	0	1	1	1	0	1	1	0	1	1	0	0	1
0.02	0.02	1.43	0.1	0	-0.01	1.4	0.08	-0.02	1.41	0.08	-0.02	-1.33	-1.43	-0.1
-0.42	-0.46	0.6	-0.62	-0.65	-0.48	0.57	-0.64	-0.67	0.57	-0.65	-0.68	-2.33	-2.38	-0.96
0.47	0.49	2.26	0.82	0.65	0.46	2.23	0.79	0.62	2.26	0.82	0.65	-0.33	-0.48	0.76
Mean														
1	1	0	1	1	1	0	0.99	1	0	1	1	0	0	1
-0.07	0	1.87	0.13	0.08	0.07	1.94	0.2	0.15	1.87	0.13	0.08	-1.74	-1.79	-0.05
-0.66	-0.62	0.78	-0.81	-0.77	-0.55	0.85	-0.74	-0.7	0.75	-0.84	-0.8	-3.07	-3.05	-1.18
0.52	0.63	2.97	1.08	0.94	0.7	3.04	1.15	1.01	2.99	1.1	0.96	-0.41	-0.53	1.08
Variance														
1	1	0	1	0.95	1	0	1	0.89	0	1	0.93	0.01	0.03	0.99
-0.51	-0.29	16.74	0.21	2.77	0.22	17.25	0.72	3.28	17.03	0.5	3.06	-16.53	-13.97	2.56
-6.64	-6.82	5.28	-9.67	-6.14	-6.31	5.79	-9.16	-5.63	5.34	-9.64	-6.14	-30.37	-27.13	-9.25
5.62	6.25	28.2	10.09	11.68	6.76	28.71	10.6	12.19	28.71	10.63	12.25	-2.69	-0.81	14.37
Standard Deviation														
0.9	1	0	1	0.92	0.98	0	0.97	0.53	0	1	0.86	0.03	0.09	0.99
-0.25	-0.07	1.72	0.05	0.35	0.18	1.97	0.31	0.6	1.79	0.12	0.42	-1.66	-1.37	0.29
-0.95	-0.81	0.41	-1.07	-0.67	-0.56	0.67	-0.82	-0.41	0.46	-1.03	-0.63	-3.24	-2.87	-1.05
0.44	0.67	3.02	1.18	1.36	0.93	3.27	1.43	1.61	3.12	1.27	1.46	-0.09	0.13	1.64
Alves Low Risk														
0.98	1	0	1	1	1	0.01	1	0.95	0	1	0.99	0.04	0	0.99
-0.02	-0.01	-0.23	-0.02	0.02	0.01	-0.21	0	0.04	-0.22	-0.01	0.03	0.21	0.25	0.04
-0.11	-0.11	-0.4	-0.16	-0.11	-0.08	-0.38	-0.14	-0.09	-0.39	-0.16	-0.11	0.01	0.05	-0.14
0.07	0.09	-0.06	0.13	0.15	0.11	-0.04	0.15	0.17	-0.05	0.14	0.16	0.41	0.44	0.21
Alves Very High Risk														
1	1	0	0.96	1	1	0	0.86	1	0	0.96	1	0	0	0.97
-0.01	0	0.25	0.03	0	0.01	0.27	0.05	0.01	0.25	0.03	0	-0.22	-0.26	-0.04
-0.08	-0.08	0.12	-0.08	-0.11	-0.06	0.13	-0.07	-0.1	0.12	-0.08	-0.11	-0.38	-0.41	-0.17
0.06	0.08	0.39	0.15	0.1	0.09	0.4	0.16	0.11	0.39	0.15	0.1	-0.06	-0.11	0.1

B-A	C-A	D-A	E-A	F-A	C-B	D-B	E-B	F-B	D-C	E-C	F-C	E-D	F-D	F-E
pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl
dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif
lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob
upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb
Number Of Private Methods (NOPM)														
Percentile High Risk														
0.86	0.83	0.9	1	0.22	0.21	0.56	1	0.03	1	0.79	0.78	0.81	0.99	0.28
-0.03	0.03	0.05	-0.02	0.09	0.07	0.08	0.01	0.12	0.02	-0.06	0.05	-0.08	0.03	0.11
-0.11	-0.05	-0.09	-0.15	-0.03	-0.02	-0.06	-0.12	0.01	-0.13	-0.18	-0.06	-0.25	-0.13	-0.04
0.05	0.12	0.2	0.1	0.2	0.15	0.23	0.13	0.23	0.17	0.07	0.17	0.1	0.2	0.26
Alves Low Risk														
0.97	0.99	0.06	1	0.81	0.76	0.17	0.93	0.98	0.03	1	0.57	0.09	0.6	0.76
-0.01	0.01	-0.08	0.01	-0.03	0.02	-0.07	0.02	-0.02	-0.09	0	-0.04	0.09	0.05	-0.04
-0.06	-0.04	-0.16	-0.06	-0.09	-0.02	-0.15	-0.05	-0.08	-0.17	-0.07	-0.1	-0.01	-0.04	-0.12
0.03	0.06	0	0.08	0.04	0.07	0.01	0.09	0.05	-0.01	0.07	0.03	0.19	0.15	0.05
Alves Medium Risk														
0.94	1	0.02	1	0.71	0.88	0.09	0.98	0.97	0.02	1	0.63	0.08	0.48	0.84
0.01	0	0.07	0	0.02	-0.01	0.06	-0.01	0.01	0.07	0	0.03	-0.07	-0.04	0.03
-0.02	-0.04	0.01	-0.05	-0.02	-0.05	0	-0.07	-0.03	0.01	-0.05	-0.02	-0.14	-0.11	-0.04
0.04	0.03	0.13	0.05	0.07	0.02	0.12	0.04	0.06	0.13	0.05	0.07	0	0.03	0.09
Type Lines of Code (TLOC)														
Minimum														
0.94	0.83	0.84	0.14	0.96	1	0.98	0.03	0.66	1	0.02	0.54	0.07	0.59	0.72
-0.79	-1.09	-1.86	3.33	1.04	-0.3	-1.07	4.11	1.83	-0.77	4.42	2.13	5.19	2.9	-2.28
-3.2	-3.66	-6.38	-0.57	-2.47	-2.88	-5.59	0.22	-1.69	-5.38	0.42	-1.5	-0.27	-2.29	-6.94
1.63	1.49	2.67	7.23	4.56	2.28	3.45	8.01	5.35	3.84	8.41	5.76	10.65	8.09	2.38
Percentile Medium Risk														
0.58	0.95	0.54	0.89	1	0.15	0.11	0.29	0.95	0.86	1	0.95	0.99	0.58	0.88
-0.04	0.02	0.07	0.04	-0.01	0.06	0.11	0.07	0.03	0.05	0.02	-0.03	-0.03	-0.08	-0.05
-0.1	-0.05	-0.05	-0.06	-0.1	-0.01	-0.01	-0.03	-0.06	-0.07	-0.09	-0.12	-0.18	-0.21	-0.17
0.03	0.09	0.19	0.14	0.09	0.13	0.23	0.18	0.12	0.17	0.12	0.07	0.11	0.06	0.08
Percentile High Risk														
1	0.41	0.65	0.87	0.62	0.43	0.66	0.86	0.61	1	0.19	0.05	0.31	0.17	1
0	-0.03	-0.04	0.03	0.03	-0.03	-0.04	0.03	0.03	-0.01	0.06	0.06	0.07	0.08	0.01
-0.04	-0.08	-0.12	-0.04	-0.03	-0.07	-0.12	-0.04	-0.03	-0.09	-0.01	0	-0.03	-0.02	-0.08
0.04	0.02	0.04	0.1	0.1	0.02	0.04	0.1	0.1	0.07	0.13	0.13	0.17	0.17	0.09

B-A	**C-A**	**D-A**	**E-A**	**F-A**	**C-B**	**D-B**	**E-B**	**F-B**	**D-C**	**E-C**	**F-C**	**E-D**	**F-D**	**F-E**
pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl
dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif
lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob
upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb
Number of Children (NC)														
Percentile High Risk														
1	1	0.04	0.62	0.99	1	0.03	0.66	0.99	0.04	0.67	0.99	0.01	0.03	0.96
-0.01	-0.01	0.5	-0.22	-0.08	0	0.51	-0.22	-0.07	0.5	-0.22	-0.07	-0.72	-0.58	0.15
-0.27	-0.28	0.02	-0.64	-0.45	-0.27	0.03	-0.63	-0.44	0.01	-0.64	-0.46	-1.3	-1.13	-0.35
0.25	0.27	0.98	0.19	0.29	0.28	0.99	0.2	0.3	0.99	0.21	0.31	-0.14	-0.03	0.64
Percentile Very High Risk														
1	1	0.04	0.62	0.99	1	0.03	0.66	0.99	0.04	0.67	0.99	0.01	0.03	0.96
0.01	0.01	-0.5	0.22	0.08	0	-0.51	0.22	0.07	-0.5	0.22	0.07	0.72	0.58	-0.15
-0.25	-0.27	-0.98	-0.19	-0.29	-0.28	-0.99	-0.2	-0.3	-0.99	-0.21	-0.31	0.14	0.03	-0.64
0.27	0.28	-0.02	0.64	0.45	0.27	-0.03	0.63	0.44	-0.01	0.64	0.46	1.3	1.13	0.35
Alves Low Risk														
0.23	0.62	0.01	1	0.99	0.99	0.26	0.73	0.93	0.16	0.91	0.99	0.07	0.12	1
-0.06	-0.04	-0.16	0	-0.02	0.01	-0.1	0.06	0.03	-0.12	0.04	0.02	0.16	0.14	-0.02
-0.13	-0.12	-0.3	-0.12	-0.13	-0.06	-0.24	-0.06	-0.07	-0.26	-0.08	-0.09	-0.01	-0.02	-0.16
0.02	0.04	-0.02	0.12	0.08	0.09	0.03	0.18	0.14	0.02	0.16	0.13	0.33	0.3	0.12
Alves Medium Risk														
0.64	1	0.01	0.92	0.36	0.49	0.08	0.38	0.94	0.01	0.97	0.27	0.01	0.52	0.2
0.02	0	0.1	-0.02	0.04	-0.03	0.07	-0.05	0.02	0.1	-0.02	0.05	-0.12	-0.05	0.06
-0.02	-0.05	0.02	-0.09	-0.02	-0.07	-0.01	-0.11	-0.04	0.02	-0.09	-0.02	-0.21	-0.14	-0.02
0.06	0.04	0.17	0.04	0.1	0.02	0.15	0.02	0.08	0.18	0.05	0.11	-0.02	0.04	0.15
Depth of Inheritance Tree (DIT)														
Percentile Low Risk														
1	0.99	0.01	0.55	1	1	0.01	0.65	1	0.01	0.81	1	0	0.02	0.87
-0.02	-0.05	0.47	-0.21	-0.04	-0.03	0.49	-0.19	-0.02	0.52	-0.16	0.01	-0.68	-0.51	0.16
-0.24	-0.28	0.06	-0.56	-0.36	-0.26	0.08	-0.54	-0.34	0.1	-0.52	-0.32	-1.17	-0.99	-0.26
0.2	0.18	0.88	0.15	0.28	0.2	0.9	0.17	0.3	0.94	0.21	0.34	-0.18	-0.04	0.59
Percentile Very High Risk														
1	1	0.02	0.61	1	1	0.01	0.71	1	0.02	0.7	1	0	0.03	0.92
0.02	0.01	-0.53	0.22	0.05	-0.01	-0.55	0.2	0.03	-0.54	0.21	0.04	0.75	0.58	-0.17
-0.23	-0.25	-1	-0.18	-0.31	-0.28	-1.02	-0.21	-0.34	-1.02	-0.21	-0.34	0.18	0.04	-0.66
0.27	0.28	-0.06	0.63	0.42	0.26	-0.08	0.61	0.4	-0.06	0.63	0.42	1.32	1.12	0.32
Alves Medium Risk														
0.37	0.05	0.27	0.09	1	0.91	0.88	0.69	0.72	1	0.97	0.29	1	0.39	0.21
0.03	0.04	0.05	0.06	0	0.01	0.03	0.03	-0.03	0.01	0.02	-0.04	0	-0.05	-0.06
-0.01	0	-0.02	0	-0.06	-0.03	-0.04	-0.03	-0.08	-0.06	-0.05	-0.1	-0.08	-0.14	-0.13
0.06	0.08	0.12	0.12	0.05	0.06	0.1	0.09	0.03	0.09	0.08	0.02	0.09	0.03	0.02

B-A	C-A	D-A	E-A	F-A	C-B	D-B	E-B	F-B	D-C	E-C	F-C	E-D	F-D	F-E
pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl
dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif
lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob
upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb
Lack of Cohesion Of Methods (LCOM)														
Percentile Medium Risk														
0.99	0.97	0.04	0.92	0.94	0.8	0.02	0.99	0.79	0.14	0.7	1	0.03	0.38	0.66
-0.01	0.01	0.09	-0.02	0.02	0.02	0.1	-0.02	0.03	0.07	-0.04	0.01	-0.11	-0.07	0.05
-0.05	-0.04	0	-0.1	-0.05	-0.03	0.01	-0.09	-0.04	-0.01	-0.11	-0.06	-0.21	-0.16	-0.04
0.04	0.06	0.17	0.05	0.09	0.07	0.18	0.06	0.1	0.16	0.04	0.08	-0.01	0.03	0.13
Alves High Risk														
0.73	0.98	0.07	0.67	1	0.34	0.01	0.99	0.94	0.2	0.4	0.99	0.01	0.15	0.83
-0.02	0.01	0.09	-0.04	0	0.04	0.11	-0.02	0.02	0.08	-0.05	-0.01	-0.13	-0.09	0.04
-0.07	-0.04	0	-0.12	-0.07	-0.02	0.02	-0.1	-0.05	-0.02	-0.14	-0.09	-0.24	-0.2	-0.06
0.03	0.07	0.18	0.04	0.07	0.09	0.21	0.06	0.1	0.17	0.03	0.06	-0.02	0.02	0.14
Fan-in (FANIN)														
Maximum														
0.76	0.47	0.55	0.73	0.96	0.99	0.17	0.99	0.5	0.09	1	0.3	0.17	0.95	0.47
2.33	3.35	-5.45	3.93	-2.09	1.02	-7.79	1.6	-4.42	-8.8	0.58	-5.44	9.39	3.36	-6.02
-2.7	-2.02	-14.87	-4.18	-9.41	-4.35	-17.2	-6.52	-11.74	-18.4	-7.74	-12.99	-1.98	-7.45	-15.72
7.37	8.71	3.96	12.05	5.23	6.38	1.63	9.72	2.9	0.8	8.91	2.11	20.75	14.17	3.68
Variance														
0.74	0.44	0.85	0.51	0.96	0.99	0.4	0.95	0.47	0.25	1	0.27	0.24	1	0.29
2.59	3.73	-4.17	5.29	-2.35	1.14	-6.76	2.7	-4.94	-7.9	1.56	-6.08	9.46	1.82	-7.64
-2.87	-2.08	-14.37	-3.51	-10.29	-4.67	-16.96	-6.1	-12.87	-18.3	-7.47	-14.27	-2.87	-9.91	-18.16
8.05	9.55	6.04	14.09	5.59	6.96	3.45	11.5	3	2.51	10.58	2.11	21.78	13.54	2.88
Standard Deviation														
0.46	0.4	0.85	0.63	0.96	1	0.28	1	0.31	0.24	1	0.27	0.31	1	0.4
0.48	0.53	-0.57	0.65	-0.31	0.06	-1.05	0.18	-0.79	-1.1	0.12	-0.84	1.22	0.26	-0.97
-0.28	-0.27	-1.99	-0.57	-1.41	-0.75	-2.46	-1.04	-1.89	-2.55	-1.13	-1.98	-0.49	-1.37	-2.42
1.23	1.34	0.85	1.87	0.79	0.86	0.37	1.4	0.31	0.34	1.37	0.29	2.93	1.88	0.49
Percentile High Risk														
0.35	0.92	0.81	0.52	0.7	0.95	0.19	1	0.06	0.47	0.9	0.29	0.21	1	0.12
-0.05	-0.03	0.06	-0.07	0.05	0.02	0.11	-0.02	0.1	0.09	-0.04	0.08	-0.13	-0.01	0.12
-0.12	-0.1	-0.08	-0.19	-0.05	-0.05	-0.03	-0.14	0	-0.05	-0.16	-0.03	-0.29	-0.16	-0.02
0.02	0.05	0.2	0.05	0.16	0.1	0.25	0.1	0.21	0.23	0.08	0.19	0.03	0.15	0.26
Alves High Risk														
1	0.99	0	0.77	1	0.94	0	0.87	1	0	0.54	0.95	0	0	0.97
0	0.01	0.08	-0.02	-0.01	0.01	0.09	-0.02	0	0.08	-0.03	-0.01	-0.1	-0.09	0.02
-0.03	-0.03	0.03	-0.07	-0.05	-0.02	0.03	-0.07	-0.05	0.02	-0.08	-0.06	-0.17	-0.15	-0.04
0.03	0.04	0.14	0.03	0.04	0.04	0.14	0.03	0.04	0.13	0.02	0.03	-0.04	-0.02	0.07

Table 4.9: Tukey differences for TA teams Categorisation C on all data

	B-A	C-A	D-A	E-A	F-A	G-A	I-A	C-B	D-B	E-B	F-B	G-B	H-B	I-B	D-C	E-C	F-C	G-C	H-C	I-C	E-D	F-D	G-D	H-D	I-D	F-E	G-E	H-E	I-E	G-F	H-F	I-F	H-G	I-G	I-H
	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb
Resultant Variable																																			
OO Standard																																			
pvl	0.28	0	0.73	0.08	0.8			0.37	0.02	0.92	1				0	0.99	0.41				0.01	0.2				0.86									
dif	-0.15	-0.3	0.12	-0.24	-0.12			-0.14	0.28	-0.09	0.03				0.42	0.06	0.18				-0.36	-0.24				0.12									
lob	-0.36	-0.51	-0.13	-0.5	-0.39			-0.35	0.03	-0.34	-0.23				0.16	-0.21	-0.09				-0.66	-0.54				-0.18									
upb	0.05	-0.08	0.38	0.02	0.15			0.07	0.52	0.17	0.3				0.68	0.32	0.45				-0.07	0.06				0.43									
Coding Standard																																			
pvl	0.33	0	0.05	0.95	0.97			0.12	0.81	0.12	0.97				0.95	0	0.07				0.02	0.51				0.72									
dif	-0.15	-0.33	-0.25	0.08	-0.07			-0.19	-0.11	0.22	0.07				0.08	0.41	0.26				0.33	0.18				-0.15									
lob	-0.35	-0.55	-0.51	-0.18	-0.34			-0.4	-0.36	-0.03	-0.19				-0.18	0.15	-0.01				0.04	-0.12				-0.46									
upb	0.06	-0.11	0	0.34	0.2			0.03	0.14	0.48	0.34				0.34	0.67	0.53				0.63	0.49				0.16									
Total																																			
Number of Methods																																			
pvl	0.81	0.93	0.95	0.9	1	0.22	1	1	0.3	0.22	0.95	0.71	0.17	1	0.44	0.34	0.98	0.64	0.21	1	1	1	0.06	0.86	1	0.99	0.05	0.9	1	0.33	0.59	1	0.03	1	0.94
dif	14.99	12.62	-15.67	-18.4	-1.89	48.2	9.78	-2.37	-30.66	-33.39	-16.88	33.21	-74.22	-5.22	-28.29	-31.02	-14.51	35.59	-71.84	-2.84	-2.73	13.78	63.87	-43.56	25.44	16.51	66.61	-40.82	28.18	50.1	-57.33	11.67	-107.43	-38.43	69
lob	-14.45	-17.95	-56.23	-59.82	-45.32	-11.26	-138.88	-32.65	-71.01	-74.61	-60.11	-26.1	-161.64	-153.81	-69.46	-73.04	-58.51	-24.3	-159.65	-151.67	-52.5	-37.67	-1.68	-135.33	-125.75	-35.62	0.52	-132.98	-123.25	-17.27	-150.41	-140.32	-208.98	-195.75	-100.93
upb	44.43	43.18	24.89	23.02	41.53	107.67	158.43	27.9	9.69	7.82	26.34	92.53	13.21	143.38	12.89	11.01	29.49	95.47	15.97	145.99	47.04	65.23	129.42	48.22	176.64	68.64	132.69	51.33	179.61	117.46	35.74	163.66	-5.88	118.9	238.93
Number of Types																																			
pvl	0.85	0.81	0.84	0.79	1	0.39	1	1	0.18	0.15	0.89	0.86	0.09	1	0.16	0.13	0.86	0.9	0.08	1	1	0.99	0.08	0.79	0.99	0.99	0.07	0.82	0.99	0.4	0.48	1	0.02	1	0.75
dif	2.78	3.04	-3.85	-4.18	-1.05	8.24	4.82	0.27	-6.63	-6.96	-3.83	5.47	-15.96	2.04	-6.89	-7.23	-4.09	5.2	-16.23	1.77	-0.33	2.8	12.1	-9.33	8.67	3.13	12.43	-9	9	9.3	-12.13	5.87	-21.43	-3.43	18
lob	-2.98	-2.93	-11.78	-12.28	-9.54	-3.38	-24.24	-5.65	-14.51	-15.02	-12.28	-6.13	-33.05	-27.01	-14.94	-15.44	-12.7	-6.51	-33.39	-27.32	-10.06	-7.26	-0.72	-27.27	-20.89	-7.06	-0.49	-27.02	-20.6	-3.87	-30.33	-23.85	-41.28	-34.18	-15.22
upb	8.53	9.02	4.08	3.91	7.44	19.87	33.88	6.19	1.26	1.1	4.62	17.06	1.13	31.09	1.16	0.99	4.51	16.91	0.94	30.87	9.4	12.86	24.91	8.61	38.22	13.32	25.35	9.02	38.6	22.46	6.06	35.58	-1.58	27.33	51.22
Method Lines of Code (MLOC)																																			
Minimum																																			
pvl	0.99	1	0.94	0.82	0.68	0.08	0.96	0.97	1	0.99	0.97	0.02	0.02	0.9	0.9	0.75	0.6	0.11	0.07	0.97	1	1	0.02	0.02	0.85	1	0.01	0.01	0.81	0.01	0.01	0.76	0.99	1	1
dif	-0.21	0.04	-0.37	-0.46	-0.56	1.24	1.24	0.25	-0.16	-0.25	-0.35	1.45	2.12	1.45	-0.41	-0.5	-0.6	1.2	1.87	1.2	-0.09	-0.19	1.61	2.28	1.61	-0.09	1.71	2.37	1.71	1.8	2.47	1.8	0.67	0	-0.67
lob	-0.86	-0.64	-1.26	-1.38	-1.52	-0.07	-2.05	-0.42	-1.05	-1.17	-1.31	0.14	0.18	-1.84	-1.32	-1.43	-1.57	-0.12	-0.07	-2.09	-1.2	-1.33	0.16	0.24	-1.74	-1.25	0.24	0.33	-1.65	0.31	0.4	-1.57	-1.58	-3.49	-4.43
upb	0.45	0.72	0.53	0.46	0.41	2.56	4.54	0.92	0.73	0.66	0.61	2.77	4.05	4.74	0.51	0.43	0.38	2.53	3.82	4.5	1.01	0.95	3.06	4.31	4.96	1.06	3.17	4.41	5.06	3.29	4.53	5.17	2.92	3.49	3.1
Cyclomatic Complexity (CC)																																			
Maximum																																			
pvl	1	0.86	0.79	0.94	0.63	1	0.99	0.65	0.62	0.83	0.77	1	0.57	0.98	1	1	0.1	1	0.92	1	1	0.1	1	0.98	1	0.2	1	0.96	1	0.69	0.19	0.85	0.95	1	1
dif	0.54	-2.35	-3.42	-2.75	4.24	-1.98	-7.69	-2.89	-3.96	-3.29	3.7	-2.52	-8.9	-8.24	-1.06	-0.4	6.59	0.37	-6.01	-5.34	0.66	7.66	1.44	-4.94	-4.28	6.99	0.77	-5.61	-4.94	-6.22	-12.6	-11.93	-6.38	-5.71	0.67
lob	-4.24	-7.32	-10.01	-9.48	-2.82	-11.64	-31.85	-7.81	-10.51	-9.99	-3.33	-12.16	-23.11	-32.38	-7.75	-7.23	-0.56	-9.36	-20.28	-29.52	-7.42	-0.7	-9.22	-19.86	-28.85	-1.48	-9.97	-20.58	-29.55	-17.17	-27.72	-36.63	-22.88	-31.28	-26.95
upb	5.33	2.61	3.17	3.98	11.3	7.68	16.46	2.03	2.6	3.4	10.72	7.12	5.3	15.91	5.63	6.43	13.74	10.1	8.26	18.84	8.75	16.02	12.09	9.97	20.29	15.46	11.51	9.37	19.67	4.73	2.52	12.76	10.12	19.85	28.28
Percentile High Risk																																			
pvl	0.96	1	0.19	1	1	1	0.99	0.96	0.71	1	0.98	1	0.08	1	0.21	1	1	1	0.22	0.99	0.69	0.36	0.89	0.01	1	1	1	0.19	1	1	0.36	0.99	0.33	1	0.46
dif	0.01	0	0.04	0.01	0	0	0.05	-0.01	0.02	-0.01	-0.02	-0.01	-0.09	0.04	0.04	0.01	0	0	-0.08	0.05	-0.03	-0.04	-0.03	-0.11	0.01	-0.01	0	-0.08	0.04	0.01	-0.07	0.05	-0.08	0.04	0.13
lob	-0.02	-0.03	-0.01	-0.04	-0.05	-0.06	-0.11	-0.05	-0.02	-0.05	-0.06	-0.07	-0.19	-0.13	-0.01	-0.04	-0.05	-0.06	-0.17	-0.11	-0.09	-0.1	-0.1	-0.21	-0.15	-0.07	-0.07	-0.18	-0.12	-0.07	-0.18	-0.11	-0.19	-0.13	-0.06
upb	0.04	0.03	0.08	0.05	0.04	0.07	0.21	0.02	0.07	0.04	0.03	0.06	0.01	0.2	0.08	0.05	0.04	0.07	0.02	0.21	0.02	0.02	0.04	-0.01	0.18	0.05	0.07	0.02	0.21	0.08	0.03	0.22	0.03	0.22	0.31
Number Of Fields (NOF)																																			
Percentile Medium Risk																																			
pvl	1	0.99	0.73	0.99	0.41	0.91	1	1	0.94	1	0.69	0.75	0	1	0.99	1	0.85	0.64	0	1	1	1	0.34	0	1	0.98	0.64	0	1	0.18	0	1	0.04	0.98	0.08
dif	-0.01	-0.02	-0.06	-0.03	-0.08	0.07	-0.07	-0.01	-0.04	-0.02	-0.06	0.08	0.35	-0.06	-0.03	-0.01	-0.05	0.09	0.35	-0.05	0.02	-0.02	0.12	0.39	-0.01	-0.04	0.1	0.36	-0.04	0.14	0.41	0.01	0.26	-0.14	-0.4
lob	-0.09	-0.1	-0.16	-0.14	-0.19	-0.08	-0.44	-0.09	-0.14	-0.12	-0.17	-0.07	0.13	-0.43	-0.14	-0.11	-0.16	-0.06	0.13	-0.42	-0.1	-0.15	-0.04	0.16	-0.4	-0.17	-0.07	0.13	-0.42	-0.03	0.17	-0.38	0.01	-0.53	-0.83
upb	0.06	0.05	0.05	0.07	0.03	0.22	0.3	0.07	0.06	0.09	0.05	0.23	0.57	0.32	0.07	0.1	0.06	0.24	0.58	0.33	0.15	0.11	0.29	0.62	0.37	0.09	0.26	0.6	0.34	0.31	0.64	0.39	0.52	0.26	0.03
Percentile High Risk																																			
pvl	1	0.19	0.98	1	0.99	0.5	1	0.11	0.95	1	0.98	0.42	0.6	1	0.99	0.78	0.99	1	0.1	0.95	1	1	0.95	0.31	0.99	1	0.75	0.54	1	0.94	0.36	0.99	0.09	0.88	1
dif	0	0.04	0.02	0	0.01	0.07	-0.06	0.05	0.03	0.01	0.02	0.07	-0.09	-0.05	-0.02	-0.04	-0.02	0.02	-0.14	-0.1	-0.02	0	0.04	-0.12	-0.08	0.01	0.06	-0.1	-0.06	0.05	-0.12	-0.08	-0.16	-0.13	0.04
lob	-0.05	-0.01	-0.05	-0.07	-0.05	-0.03	-0.32	0	-0.04	-0.06	-0.05	-0.03	-0.24	-0.31	-0.09	-0.11	-0.1	-0.08	-0.29	-0.36	-0.1	-0.09	-0.07	-0.28	-0.34	-0.08	-0.05	-0.26	-0.33	-0.07	-0.28	-0.34	-0.34	-0.4	-0.26
upb	0.05	0.1	0.09	0.08	0.1	0.17	0.2	0.1	0.1	0.08	0.1	0.17	0.06	0.2	0.05	0.03	0.05	0.13	0.01	0.16	0.07	0.09	0.16	0.04	0.18	0.1	0.18	0.06	0.2	0.16	0.04	0.18	0.01	0.15	0.33
Number Of Private Fields (NOPF)																																			
Mean																																			
pvl	1	1	0.01	1	1	1	1	1	0.01	1	1	0.99	1	1	0.01	1	1	1	1	1	0.06	0.23	0.82	0.66	0.98	1	1	1	1	1	1	1	1	1	1
dif	-0.05	0	0.84	-0.03	0.1	0.25	-0.09	0.05	0.89	0.02	0.15	0.3	-0.08	-0.04	0.85	-0.03	0.1	0.25	-0.13	-0.09	-0.88	-0.75	-0.6	-0.97	-0.94	0.13	0.28	-0.1	-0.06	0.15	-0.23	-0.19	-0.38	-0.34	0.04
lob	-0.58	-0.56	0.11	-0.78	-0.69	-0.83	-2.79	-0.5	0.16	-0.73	-0.64	-0.78	-1.67	-2.74	0.1	-0.79	-0.7	-0.83	-1.72	-2.79	-1.78	-1.68	-1.78	-2.64	-3.68	-0.82	-0.92	-1.77	-2.81	-1.07	-1.91	-2.94	-2.22	-3.19	-3.04
upb	0.48	0.55	1.58	0.72	0.88	1.32	2.6	0.59	1.62	0.76	0.93	1.37	1.5	2.65	1.59	0.73	0.9	1.34	1.46	2.61	0.03	0.18	0.59	0.69	1.8	1.07	1.48	1.57	2.68	1.37	1.46	2.57	1.46	2.51	3.12
Alves Very High Risk																																			
pvl	1	1	0.02	1	1	0.97	1	1	0	1	0.99	0.88	1	1	0.02	1	1	0.97	1	1	0.15	0.32	0.96	0.4	0.92	1	0.99	1	1	1	0.99	1	0.95	1	1
dif	-0.02	0	0.11	0.01	0.02	0.05	-0.05	0.02	0.13	0.02	0.03	0.07	-0.03	-0.03	0.12	0.01	0.02	0.05	-0.05	-0.05	-0.11	-0.1	-0.06	-0.16	-0.16	0.01	0.05	-0.05	-0.05	0.03	-0.07	-0.07	-0.1	-0.1	0
lob	-0.09	-0.08	0.01	-0.1	-0.09	-0.1	-0.42	-0.06	0.03	-0.08	-0.07	-0.08	-0.25	-0.4	0.01	-0.1	-0.09	-0.1	-0.27	-0.42	-0.23	-0.23	-0.23	-0.39	-0.54	-0.12	-0.12	-0.28	-0.43	-0.13	-0.3	-0.45	-0.35	-0.49	-0.43
upb	0.06	0.07	0.22	0.11	0.13	0.2	0.32	0.09	0.23	0.13	0.14	0.22	0.19	0.34	0.22	0.11	0.13	0.2	0.17	0.33	0.02	0.03	0.1	0.07	0.22	0.14	0.21	0.18	0.33	0.2	0.17	0.32	0.15	0.29	0.43

B-A	C-A	D-A	E-A	F-A	G-A	I-A	C-B	D-B	E-B	F-B	G-B	H-B	I-B	D-C	E-C	F-C	G-C	H-C	I-C	E-D	F-D	G-D	H-D	I-D	F-E	G-E	H-E	I-E	G-F	H-F	I-F	H-G	I-G	I-H
pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl
dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif
lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob
upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb
Number Of Private Methods (NOPM)																																		
Percentile High Risk																																		
0.99	1	1	1	1	1	1	0.8	0.86	0.99	0.99	1	0	1	1	1	1	0.98	0	1	1	1	0.97	0	1	1	1	0	1	1	0	1	0	1	0.16
-0.02	0.02	0.03	0.01	0.01	-0.03	-0.01	0.04	0.05	0.03	0.03	-0.01	0.37	0.01	0.01	-0.01	-0.01	-0.05	0.33	-0.03	-0.02	-0.02	-0.06	0.32	-0.04	0	-0.04	0.34	-0.02	-0.04	0.34	-0.02	0.38	0.02	-0.36
-0.09	-0.06	-0.07	-0.09	-0.1	-0.18	-0.37	-0.04	-0.05	-0.07	-0.07	-0.15	0.15	-0.35	-0.09	-0.11	-0.11	-0.19	0.11	-0.39	-0.14	-0.14	-0.22	0.09	-0.41	-0.13	-0.2	0.11	-0.39	-0.21	0.11	-0.39	0.13	-0.37	-0.77
0.05	0.09	0.13	0.11	0.12	0.12	0.35	0.11	0.15	0.13	0.14	0.14	0.58	0.37	0.11	0.1	0.1	0.1	0.54	0.34	0.11	0.11	0.1	0.54	0.33	0.13	0.12	0.56	0.35	0.12	0.56	0.35	0.62	0.4	0.06
Alves Low Risk																																		
1	1	0.3	0.98	0.01	1	1	0.98	0.67	1	0.04	1	0.89	0.99	0.23	0.95	0	1	0.98	1	0.98	0.93	0.69	0.44	0.9	0.35	0.99	0.82	0.98	0.14	0.11	0.67	1	1	1
-0.01	0	-0.05	-0.02	-0.08	0.01	0.06	0.02	-0.04	-0.01	-0.07	0.02	0.06	0.07	-0.05	-0.03	-0.09	0.01	0.05	0.05	0.03	-0.03	0.06	0.1	0.11	-0.06	0.03	0.07	0.08	0.09	0.14	0.14	0.04	0.05	0.01
-0.06	-0.05	-0.11	-0.09	-0.15	-0.08	-0.18	-0.03	-0.1	-0.08	-0.14	-0.07	-0.08	-0.17	-0.12	-0.09	-0.16	-0.09	-0.09	-0.19	-0.05	-0.12	-0.05	-0.05	-0.14	-0.15	-0.07	-0.07	-0.16	-0.01	-0.01	-0.1	-0.12	-0.21	-0.27
0.04	0.05	0.02	0.04	-0.01	0.11	0.3	0.06	0.03	0.06	0	0.12	0.2	0.31	0.01	0.04	-0.02	0.1	0.19	0.29	0.11	0.05	0.17	0.25	0.35	0.02	0.14	0.22	0.32	0.2	0.29	0.39	0.2	0.3	0.28
Type Lines of Code (TLOC)																																		
Minimum																																		
0.99	1	0.99	0.79	1	0.01	1	1	1	0.99	1	0	0.1	1	1	0.98	1	0	0.13	1	1	1	0	0.12	1	0.94	0	0.05	1	0.03	0.29	1	1	0.94	0.96
-0.7	-0.55	-0.9	-1.7	-0.01	5.65	0.65	0.15	-0.2	-1	0.69	6.35	6.35	1.35	-0.35	-1.15	0.54	6.2	6.2	1.2	-0.8	0.89	6.56	6.56	1.56	1.69	7.35	7.35	2.35	5.67	5.67	0.67	0	-5	-5
-3.04	-2.98	-4.13	-4.99	-3.47	0.93	-11.16	-2.26	-3.41	-4.28	-2.75	1.64	-0.6	-10.46	-3.62	-4.49	-2.96	1.44	-0.78	-10.62	-4.75	-3.2	1.35	-0.74	-10.46	-2.46	2.1	0.03	-9.68	0.31	-1.73	-11.41	-8.07	-17.5	-18.51
1.64	1.88	2.32	1.59	3.44	10.38	12.47	2.56	3	2.28	4.12	11.07	13.3	13.16	2.92	2.19	4.04	10.96	13.18	13.03	3.16	4.98	11.77	13.85	13.57	5.83	12.61	14.68	14.39	11.02	13.06	12.75	8.07	7.5	8.51
Variance																																		
1	0.99	0.2	1	0.24	1	1	1	0.1	1	0.13	1	0.99	1	0.04	1	0.06	1	1	1	0.17	1	0.7	0.44	1	0.19	1	1	1	0.7	0.44	1	1	1	1
-	-	2261.9	-	2339	-	-	-	2493.8	-	2570.8	-13.14	-	49.52	2841.3	-1.14	2918.3	334.34	-	397	-	77.02	-	-	-	2919.4	335.48	-	398.14	-	-	-	-	62.66	1802.41
231.83	579.32		580.45		244.97	182.31	347.48		348.62			1752.9						1405.4		2842.4		2506.9	4246.7	2444.3			1404.2		2583.9	4323.7	2521.3	1739.7		
-	-	-	-	-	-4265	-	-	-	-	-	-	-	-	57.65	-	-56.33	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2222.0	2645.6	480.13	3380.9	596.81		10232.	2394.5	233.81	3134.9	351.46	4023.3	7663.5	9996.6		2842.3		3714.1	7342.1	9664.4	6207.2	3401.2	6938.6	10451.	12666	604.96	4132.5	7634.6	9839.3	7138.0	10616.	12796.	8605.2	10573.	9685.78
1758.3	1487.0	5004.1	2220.0	5274.8	3775.0	9867.7	1699.5	5221.4	2437.6	5493.1	3997.0	4157.7	10095.	5624.9	2840.0	5892.9	4382.8	4531.2	10458.	522.34	3555.2	1924.7	1957.6	7777.4	6443.8	4803.4	4826.0	10635.	1970.0	1968.6	7754.0	5125.7	10698.	13290.61
Standard Deviation																																		
1	0.92	0.41	1	0.55	1	1	0.97	0.3	1	0.43	1	0.86	1	0.05	1	0.09	1	0.98	1	0.35	1	0.94	0.26	1	0.44	1	0.96	1	0.96	0.29	1	0.9	1	0.99
-1.07	-6.31	13.67	-3.93	13.28	0.88	2.39	-5.24	14.74	-2.85	14.35	1.95	-19.82	3.47	19.98	2.39	19.59	7.19	-14.58	8.71	-17.6	-0.39	-12.79	-34.57	-11.28	17.2	4.81	-16.97	6.32	-12.4	-34.17	-10.88	-21.78	1.51	23.29
-15.25	-21.03	-5.86	-23.87	-7.63	-27.75	-69.18	-19.82	-4.68	-22.7	-6.46	-26.6	-61.91	-68.08	0.16	-17.85	-1.59	-21.64	-56.86	-62.95	-41.56	-25.16	-44.35	-78.75	-84.07	-7.9	-27.01	-61.34	-66.59	-44.83	-78.99	-84.06	-70.67	-74.23	-58.53
13.1	8.4	33.2	16.02	34.19	29.51	73.97	9.34	34.17	16.99	35.16	30.51	22.27	75.01	39.81	22.62	40.77	36.02	27.7	80.36	6.37	24.38	18.77	9.62	61.52	42.3	36.63	27.4	79.23	20.03	10.64	62.29	27.12	77.26	105.11
Alves Very High Risk																																		
1	1	0.42	1	0.24	1	0.99	1	0.33	1	0.18	1	1	0.99	0.4	1	0.22	1	1	0.99	0.25	1	0.57	0.81	1	0.14	1	1	0.97	0.4	0.69	1	1	0.97	0.97
0	0	0.02	-0.01	0.02	-0.01	0.03	0	0.02	-0.01	0.02	-0.01	-0.01	0.03	0.02	-0.01	0.02	-0.01	-0.01	0.03	-0.03	0	-0.03	-0.03	0.01	0.03	0	0	0.03	-0.03	-0.03	0	0	0.03	0.04
-0.02	-0.02	-0.01	-0.03	-0.01	-0.05	-0.07	-0.02	-0.01	-0.03	0	-0.04	-0.07	-0.07	-0.01	-0.03	-0.01	-0.05	-0.07	-0.07	-0.06	-0.03	-0.07	-0.09	-0.09	0	-0.04	-0.06	-0.06	-0.07	-0.09	-0.09	-0.07	-0.06	-0.07
0.02	0.02	0.04	0.02	0.05	0.03	0.12	0.02	0.04	0.02	0.05	0.03	0.04	0.12	0.04	0.02	0.05	0.03	0.04	0.12	0.01	0.04	0.02	0.03	0.1	0.06	0.04	0.05	0.13	0.01	0.03	0.1	0.06	0.13	0.15
Number of Children (NC)																																		
Maximum																																		
0.57	0.3	1	1	1	0.83	1	1	0.88	0.94	0.92	0.22	0.73	0.98	0.7	0.8	0.78	0.13	0.62	0.97	1	1	0.89	0.98	1	1	0.86	0.98	1	0.9	0.99	1	1	1	1
0.87	1.09	0	0.08	0.01	-1.39	-1.39	0.22	-0.87	-0.78	-0.85	-2.25	-2.25	-2.25	-1.09	-1.01	-1.08	-2.48	-2.48	-2.48	0.08	0.01	-1.39	-1.39	-1.39	-0.07	-1.47	-1.47	-1.47	-1.4	-1.4	-1.4	0	0	0
-0.51	-0.35	-1.9	-1.86	-2.03	-4.18	-8.37	-1.2	-2.76	-2.72	-2.88	-5.04	-6.36	-9.23	-3.02	-2.98	-3.14	-5.29	-6.6	-9.46	-2.25	-2.4	-4.47	-5.7	-8.49	-2.52	-4.57	-5.8	-8.58	-4.56	-5.77	-8.53	-4.77	-7.38	-7.98
2.25	2.52	1.9	2.03	2.05	1.4	5.59	1.64	1.03	1.15	1.17	0.53	1.85	4.72	0.84	0.97	0.99	0.33	1.64	4.51	2.42	2.43	1.69	2.92	5.71	2.38	1.63	2.85	5.64	1.76	2.97	5.73	4.77	7.38	7.98
Lack of Cohesion Of Methods (LCOM)																																		
Percentile Medium Risk																																		
1	1	0.85	0.59	0.03	0.99	1	1	0.87	0.62	0.03	0.99	1	1	0.97	0.83	0.09	0.96	1	1	1	0.8	0.71	0.94	1	0.95	0.53	0.87	0.99	0.09	0.46	0.93	1	1	1
0	0.01	0.03	0.05	0.08	-0.03	-0.03	0.01	0.03	0.04	0.08	-0.03	-0.03	-0.03	0.03	0.04	0.07	-0.04	-0.04	-0.04	0.01	0.05	-0.07	-0.07	-0.07	0.04	-0.08	-0.08	-0.08	-0.11	-0.11	-0.11	0	0	0
-0.05	-0.05	-0.04	-0.03	0	-0.14	-0.3	-0.05	-0.04	-0.03	0	-0.14	-0.19	-0.3	-0.05	-0.04	-0.01	-0.15	-0.2	-0.3	-0.08	-0.04	-0.18	-0.23	-0.33	-0.06	-0.19	-0.24	-0.35	-0.23	-0.28	-0.38	-0.18	-0.28	-0.3
0.05	0.06	0.11	0.12	0.16	0.08	0.23	0.06	0.11	0.12	0.16	0.07	0.12	0.23	0.1	0.11	0.15	0.07	0.12	0.23	0.1	0.14	0.05	0.1	0.2	0.13	0.04	0.09	0.19	0.01	0.05	0.16	0.18	0.28	0.3

	B-A	C-A	D-A	E-A	F-A	G-A	I-A	C-B	D-B	E-B	F-B	G-B	H-B	I-B	D-C	E-C	F-C	G-C	H-C	I-C	E-D	F-D	G-D	H-D	I-D	F-E	G-E	H-E	I-E	G-F	H-F	I-F	H-G	I-G	I-H
	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb
Fan-in (FANIN)																																			
Minimum																																			
	1	1	1	0.34	1	1	1	1	1	0.34	1	1	0	1	1	0.36	1	1	0	1	0.6	1	1	0	1	0.66	0.88	0	1	1	0	1	0	1	0.04
	0	0	0	0.06	0	0	0	0	0	0.06	0	0	0.33	0	0	0.06	0	0	0.33	0	0.06	0	0	0.33	0	-0.06	-0.06	0.27	-0.06	0	0.33	0	0.33	0	-0.33
	-0.06	-0.06	-0.08	-0.02	-0.08	-0.11	-0.29	-0.06	-0.08	-0.02	-0.08	-0.11	0.16	-0.29	-0.08	-0.02	-0.08	-0.12	0.16	-0.29	-0.04	-0.1	-0.13	0.16	-0.29	-0.16	-0.19	0.1	-0.35	-0.13	0.15	-0.29	0.14	-0.3	-0.66
	0.06	0.06	0.08	0.14	0.08	0.11	0.29	0.06	0.08	0.14	0.08	0.11	0.5	0.29	0.08	0.14	0.08	0.12	0.5	0.29	0.15	0.1	0.13	0.51	0.29	0.04	0.07	0.45	0.23	0.13	0.51	0.29	0.53	0.3	-0.01
Maximum																																			
	0.9	0.99	0.87	0.84	1	0.44	0.97	1	0.24	0.22	0.76	0.87	0.55	1	0.45	0.42	0.91	0.77	0.65	0.99	1	1	0.11	1	0.86	1	0.1	1	0.85	0.31	0.97	0.93	0.21	1	0.71
	2.07	1.44	-2.99	-3.18	-1.58	6.43	8.29	-0.63	-5.06	-5.25	-3.65	4.36	-8.78	6.22	-4.44	-4.63	-3.03	4.98	-8.16	6.84	-0.19	1.41	9.42	-3.72	11.28	1.6	9.61	-3.53	11.47	8.01	-5.13	9.87	-13.14	1.86	15
	-2.58	-3.38	-9.4	-9.73	-8.44	-2.96	-15.19	-5.41	-11.43	-11.76	-10.48	-5.01	-22.59	-17.26	-10.94	-11.27	-9.98	-4.47	-22.03	-16.67	-8.05	-6.72	-0.93	-18.22	-12.6	-6.63	-0.83	-18.09	-12.45	-2.63	-19.83	-14.14	-29.18	-22.99	-11.84
	6.72	6.27	3.41	3.36	5.28	15.82	31.77	4.16	1.31	1.25	3.18	13.73	5.03	29.69	2.07	2.01	3.92	14.44	5.71	30.35	7.67	9.54	19.77	10.77	35.16	9.84	20.05	11.03	35.39	18.65	9.57	33.87	2.9	26.71	41.84
Variance																																			
	0.89	0.93	0.9	0.98	1	0.04	0.99	1	0.27	0.5	0.69	0.22	0.81	1	0.33	0.56	0.74	0.22	0.83	1	1	1	0.01	1	0.93	1	0.01	1	0.95	0.02	1	0.97	0.11	1	0.9
	2.23	2.1	-3	-2.31	-1.89	10.21	7.38	-0.12	-5.23	-4.53	-4.12	7.98	-7.38	5.16	-5.1	-4.41	-3.99	8.11	-7.26	5.28	0.69	1.11	13.21	-2.16	10.38	0.41	12.52	-2.85	9.69	12.1	-3.27	9.27	-15.37	-2.83	12.54
	-2.68	-2.99	-9.76	-9.21	-9.13	0.3	-17.4	-5.17	-11.95	-11.4	-11.32	-1.9	-21.96	-19.61	-11.97	-11.41	-11.33	-1.87	-21.9	-19.53	-7.6	-7.47	2.28	-17.46	-14.82	-8.28	1.5	-18.22	-15.56	0.87	-18.78	-16.06	-32.3	-29.05	-15.78
	7.13	7.2	3.76	4.6	5.35	20.12	32.16	4.92	1.5	2.34	3.09	17.87	7.19	29.93	1.76	2.6	3.34	18.09	7.38	30.09	8.99	9.69	24.14	13.14	35.59	9.1	23.53	12.51	34.93	23.33	12.25	34.61	1.56	23.4	40.87
Standard Deviation																																			
	0.7	0.99	0.96	0.97	1	0.18	0.97	1	0.24	0.28	0.61	0.72	0.41	1	0.61	0.67	0.91	0.47	0.6	0.99	1	1	0.05	0.99	0.9	1	0.06	0.99	0.9	0.13	0.96	0.94	0.09	1	0.69
	0.41	0.23	-0.38	-0.37	-0.23	1.22	1.27	-0.18	-0.79	-0.78	-0.64	0.81	-1.5	0.87	-0.61	-0.6	-0.46	0.99	-1.32	1.04	0.01	0.15	1.6	-0.71	1.65	0.13	1.59	-0.72	1.64	1.45	-0.86	1.51	-2.31	0.06	2.36
	-0.31	-0.52	-1.37	-1.38	-1.3	-0.24	-2.37	-0.92	-1.78	-1.78	-1.7	-0.64	-3.64	-2.77	-1.62	-1.63	-1.54	-0.48	-3.47	-2.6	-1.21	-1.11	-0.01	-2.96	-2.05	-1.14	-0.03	-2.98	-2.07	-0.2	-3.13	-2.21	-4.79	-3.8	-1.8
	1.13	0.98	0.61	0.65	0.83	2.67	4.91	0.56	0.2	0.23	0.42	2.26	0.64	4.5	0.4	0.43	0.61	2.45	0.83	4.69	1.23	1.41	3.2	1.54	5.36	1.41	3.2	1.53	5.35	3.1	1.42	5.23	0.18	3.91	6.52
Percentile High Risk																																			
	0.94	1	0.84	0.73	0.99	0.61	1	0.91	0.27	0.18	0.68	0.93	0.07	1	0.9	0.81	1	0.56	0.21	1	1	1	0.16	0.67	0.96	1	0.12	0.74	0.95	0.35	0.53	0.98	0.02	1	0.53
	-0.03	0	0.05	0.06	0.03	-0.1	-0.1	0.04	0.08	0.09	0.07	-0.06	0.22	-0.07	0.05	0.06	0.03	-0.1	0.19	-0.1	0.01	-0.02	-0.15	0.14	-0.15	-0.03	-0.16	0.13	-0.16	-0.13	0.16	-0.13	0.29	0	-0.29
	-0.11	-0.08	-0.06	-0.05	-0.08	-0.25	-0.49	-0.05	-0.02	-0.02	-0.05	-0.22	-0.01	-0.46	-0.06	-0.06	-0.09	-0.26	-0.04	-0.49	-0.12	-0.15	-0.32	-0.1	-0.55	-0.16	-0.33	-0.11	-0.56	-0.31	-0.09	-0.53	0.02	-0.42	-0.74
	0.05	0.08	0.16	0.17	0.15	0.06	0.3	0.12	0.19	0.2	0.18	0.09	0.46	0.33	0.16	0.17	0.15	0.06	0.42	0.29	0.14	0.12	0.03	0.38	0.25	0.11	0.02	0.38	0.24	0.05	0.4	0.27	0.56	0.42	0.16
Alves Very High Risk																																			
	0.71	1	1	0.06	0.71	1	1	0.93	0.98	0.65	1	0.84	0.72	1	1	0.14	0.88	0.99	0.93	1	0.31	0.94	1	0.95	1	0.99	0.23	0.25	1	0.74	0.61	1	1	0.99	0.96
	0.01	0	0	0.02	0.01	-0.01	0.02	-0.01	-0.01	0.01	0	-0.02	-0.03	0.01	0	0.02	0.01	-0.01	-0.02	0.02	0.02	0.01	-0.01	-0.02	0.02	-0.01	-0.03	-0.04	-0.01	-0.02	-0.03	0	-0.01	0.03	0.04
	-0.01	-0.02	-0.02	0	-0.01	-0.04	-0.07	-0.03	-0.03	-0.01	-0.02	-0.05	-0.08	-0.08	-0.03	0	-0.01	-0.05	-0.08	-0.07	-0.01	-0.02	-0.05	-0.08	-0.08	-0.04	-0.07	-0.1	-0.1	-0.06	-0.09	-0.09	-0.07	-0.07	-0.07
	0.03	0.02	0.03	0.05	0.04	0.03	0.11	0.01	0.02	0.04	0.03	0.02	0.02	0.1	0.02	0.05	0.04	0.03	0.03	0.11	0.05	0.04	0.03	0.03	0.11	0.02	0.01	0.01	0.09	0.02	0.02	0.1	0.05	0.12	0.14
Fan-out (FANOUT)																																			
Alves High Risk																																			
	1	1	1	1	1	0	1	1	1	1	1	0	0.82	1	1	0.99	1	0	0.86	1	1	1	0.01	0.89	1	0.99	0	0.67	1	0.02	0.93	1	0.98	0.38	0.87
	0	0.01	0	-0.01	0.01	0.08	-0.03	0	0	-0.01	0.01	0.08	0.05	-0.04	0	-0.01	0	0.08	0.04	-0.04	-0.01	0	0.08	0.04	-0.04	0.02	0.09	0.06	-0.03	0.08	0.04	-0.04	-0.04	-0.12	-0.08
	-0.03	-0.03	-0.04	-0.05	-0.04	0.02	-0.19	-0.03	-0.04	-0.05	-0.04	0.02	-0.05	-0.19	-0.04	-0.06	-0.04	0.02	-0.05	-0.2	-0.07	-0.05	0.01	-0.05	-0.2	-0.04	0.02	-0.04	-0.19	0.01	-0.06	-0.2	-0.14	-0.28	-0.26
	0.03	0.04	0.05	0.04	0.05	0.15	0.12	0.03	0.05	0.03	0.05	0.14	0.14	0.12	0.04	0.03	0.05	0.14	0.14	0.12	0.04	0.06	0.15	0.14	0.12	0.07	0.16	0.15	0.13	0.15	0.14	0.12	0.07	0.05	0.1

Table 4.10: Tukey differences for TA teams Categorisation C on complete data

B-A	C-A	D-A	E-A	F-A	G-A	I-A	C-B	D-B	E-B	F-B	G-B	H-B	I-B	D-C	E-C	F-C	G-C	H-C	I-C	E-D	F-D	G-D	H-D	I-D	F-E	G-E	H-E	I-E	G-F	H-F	I-F	H-G	I-G	I-H
pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl	pvl
dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif	dif
lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob	lob
upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb	upb
Resultant Variable																																		
OO Standard																																		
0.19	0.09	1	0.17	1			1	0.57	0.82	0.81				0.39	0.94	0.62				0.27	1				0.42									
-0.19	-0.23	0.02	-0.38	-0.04			-0.04	0.21	-0.2	0.14				0.25	-0.15	0.19				-0.41	-0.07				0.34									
-0.42	-0.48	-0.35	-0.85	-0.38			-0.29	-0.16	-0.66	-0.19				-0.13	-0.63	-0.16				-0.95	-0.51				-0.19									
0.05	0.02	0.39	0.08	0.29			0.21	0.58	0.27	0.48				0.63	0.32	0.53				0.14	0.38				0.87									
Coding Standard																																		
0.28	0.01	0.16	1	1			0.76	0.89	0.69	0.8				1	0.27	0.25				0.37	0.43				1									
-0.17	-0.28	-0.3	0.06	-0.02			-0.11	-0.13	0.23	0.14				-0.02	0.35	0.26				0.37	0.28				-0.09									
-0.4	-0.53	-0.67	-0.4	-0.36			-0.36	-0.5	-0.23	-0.19				-0.4	-0.12	-0.08				-0.18	-0.16				-0.61									
0.06	-0.04	0.06	0.53	0.31			0.13	0.23	0.7	0.48				0.36	0.82	0.6				0.91	0.71				0.43									
Total																																		
Number of Methods																																		
0.89	0.59	0.32	1	1	0.42	1	1	0.06	1	1	0.87	0.15	1	0.03	1	1	0.97	0.09	1	0.68	0.44	0.02	1	0.99	1	0.95	0.58	1	0.83	0.43	1	0.04	1	0.95
14.93	21.81	-46.33	8.49	6.35	46.07	5.24	6.89	-61.26	-6.44	-8.58	31.14	-78.69	-9.69	-68.15	-13.33	-15.46	24.26	-85.58	-16.58	54.82	52.68	92.4	-17.43	51.57	-2.14	37.58	-72.25	-3.25	39.72	-70.11	-1.11	-109.83	-40.83	69
-18.38	-13.7	-108.65	-71.39	-49.72	-20.55	-149.22	-28.62	-123.58	-86.32	-64.65	-35.48	-169.92	-164.15	-131.67	-94.14	-72.87	-43.49	-177.63	-171.52	-40.86	-24.25	7.48	-122.77	-111.62	-93.87	-60.95	-188.84	-173.92	-40.73	-171.88	-162.02	-217.77	-205.71	-107.27
48.24	57.32	15.99	88.37	62.42	112.69	159.69	42.39	1.06	73.44	47.49	97.76	12.54	144.77	-4.63	67.49	41.94	92.01	6.48	138.37	150.5	129.61	177.33	87.91	214.76	89.59	136.12	44.34	167.42	120.18	31.66	159.8	-1.89	124.05	245.27
Number of Types																																		
0.89	0.18	0.36	0.99	1	0.44	1	0.93	0.07	1	0.99	0.88	0.08	1	0.01	1	0.71	1	0.02	1	0.38	0.71	0.03	1	0.93	0.99	1	0.21	1	0.63	0.44	1	0.02	1	0.76
2.88	5.69	-8.62	4.49	-0.43	8.74	4.24	2.81	-11.5	1.61	-3.31	5.86	-16.64	1.36	-14.31	-1.2	-6.12	3.05	-19.45	-1.45	13.11	8.19	17.36	-5.14	12.86	-4.92	4.25	-18.25	-0.25	9.17	-13.33	4.67	-22.5	-4.5	18
-3.5	-1.11	-20.56	-10.81	-11.17	-4.02	-25.35	-3.99	-23.44	-13.69	-14.05	-6.9	-34.12	-28.23	-26.48	-16.68	-17.12	-9.93	-37.09	-31.13	-5.22	-6.54	1.09	-25.32	-18.4	-22.49	-14.62	-40.58	-32.94	-6.24	-32.83	-26.15	-43.17	-36.08	-15.76
9.26	12.49	3.32	19.79	10.31	21.5	33.82	9.61	0.44	16.91	7.43	18.62	0.83	30.94	-2.15	14.27	4.87	16.02	-1.82	28.22	31.43	22.92	33.62	15.03	44.11	12.65	23.12	4.08	32.44	24.58	6.16	35.49	-1.83	27.08	51.76
Method Lines of Code (MLOC)																																		
Minimum																																		
1	1	0.94	0.85	0.45	0.23	0.97	1	0.99	0.95	0.74	0.1	0.05	0.94	0.99	0.94	0.71	0.13	0.06	0.95	1	1	0.09	0.04	0.85	1	0.09	0.03	0.75	0.01	0.01	0.69	0.99	1	1
-0.17	-0.14	-0.55	-0.83	-0.83	1.17	1.17	0.03	-0.38	-0.67	-0.67	1.33	2	1.33	-0.41	-0.7	-0.7	1.3	1.97	1.3	-0.29	-0.29	1.71	2.38	1.71	0	2	2.67	2	2	2.67	2	0.67	0	-0.67
-0.89	-0.91	-1.91	-2.58	-2.06	-0.29	-2.2	-0.74	-1.74	-2.41	-1.89	-0.12	0.01	-2.04	-1.8	-2.46	-1.95	-0.18	-0.04	-2.08	-2.37	-1.96	-0.14	0.08	-1.85	-2	-0.15	0.12	-1.73	0.24	0.45	-1.51	-1.69	-3.6	-4.51
0.56	0.64	0.81	0.91	0.39	2.62	4.54	0.81	0.98	1.08	0.56	2.79	3.99	4.7	0.98	1.07	0.56	2.78	3.98	4.68	1.8	1.39	3.57	4.68	5.28	2	4.15	5.21	5.73	3.76	4.89	5.51	3.02	3.6	3.18
Mean																																		
0.75	0.22	0.81	0.65	0.77	1	0.99	0.99	0.25	0.97	1	1	1	1	0.08	1	1	1	1	1	0.21	0.25	0.72	0.51	0.89	1	0.99	1	1	1	1	1	1	1	1
-0.71	-1.14	1.24	-1.86	-1.17	-0.62	-1.66	-0.43	1.95	-1.15	-0.46	0.09	-0.78	-0.95	2.39	-0.72	-0.03	0.52	-0.34	-0.52	-3.1	-2.41	-1.87	-2.73	-2.91	0.69	1.24	0.37	0.2	0.55	-0.32	-0.49	-0.86	-1.04	-0.18
-2.03	-2.55	-1.23	-5.03	-3.4	-3.27	-7.8	-1.84	-0.52	-4.32	-2.69	-2.56	-4.4	-7.09	-0.14	-3.93	-2.31	-2.17	-4	-6.67	-6.9	-5.47	-5.24	-6.91	-9.39	-2.95	-2.68	-4.26	-6.58	-2.65	-4.36	-6.88	-5.15	-7.59	-7.18
0.61	0.27	3.72	1.32	1.06	2.02	4.47	0.98	4.43	2.02	1.77	2.73	2.85	5.18	4.91	2.49	2.25	3.21	3.31	5.63	0.7	0.64	1.51	1.45	3.58	4.33	5.15	5	6.97	3.74	3.72	5.9	3.42	5.51	6.83
Percentile Low Risk																																		
0.96	0.62	0.97	0.9	0.48	0.86	1	1	0.71	0.99	0.89	0.49	0.84	1	0.43	1	0.99	0.26	0.95	1	0.61	0.27	1	0.31	1	1	0.45	1	1	0.16	1	1	0.21	1	0.99
0.02	0.04	-0.04	0.06	0.07	-0.06	0.01	0.02	-0.06	0.04	0.04	-0.08	0.08	-0.01	-0.08	0.03	0.03	-0.09	0.06	-0.03	0.1	0.11	-0.02	0.14	0.05	0	-0.12	0.04	-0.05	-0.12	0.03	-0.06	0.16	0.07	-0.09
-0.04	-0.03	-0.15	-0.08	-0.03	-0.18	-0.27	-0.05	-0.17	-0.1	-0.06	-0.2	-0.08	-0.29	-0.19	-0.12	-0.07	-0.22	-0.1	-0.31	-0.07	-0.03	-0.17	-0.05	-0.24	-0.16	-0.3	-0.17	-0.36	-0.27	-0.15	-0.34	-0.04	-0.23	-0.41
0.08	0.1	0.07	0.21	0.17	0.06	0.29	0.08	0.05	0.18	0.15	0.04	0.24	0.26	0.04	0.17	0.13	0.03	0.23	0.25	0.27	0.24	0.14	0.33	0.34	0.17	0.06	0.25	0.25	0.02	0.22	0.23	0.35	0.36	0.22
Percentile Medium Risk																																		
1	0.92	0.94	1	0.49	0.79	1	1	0.74	1	0.82	0.51	0.75	1	0.56	1	0.95	0.36	0.86	1	0.98	0.23	1	0.25	1	0.98	0.93	0.89	1	0.13	1	0.98	0.16	1	0.93
-0.01	-0.02	0.03	-0.01	-0.04	0.04	0.02	-0.01	0.04	0	-0.03	0.05	-0.05	0.03	0.04	0.01	-0.02	0.05	-0.05	0.03	-0.03	-0.07	0.01	-0.09	-0.01	-0.03	0.04	-0.06	0.02	0.08	-0.02	0.06	-0.1	-0.02	0.08
-0.05	-0.06	-0.04	-0.09	-0.1	-0.04	-0.15	-0.05	-0.03	-0.08	-0.09	-0.03	-0.15	-0.14	-0.03	-0.08	-0.09	-0.02	-0.15	-0.14	-0.14	-0.15	-0.08	-0.21	-0.19	-0.13	-0.06	-0.18	-0.16	-0.01	-0.13	-0.12	-0.22	-0.2	-0.11
0.03	0.02	0.1	0.08	0.02	0.11	0.18	0.03	0.11	0.09	0.03	0.12	0.05	0.19	0.11	0.1	0.04	0.13	0.05	0.2	0.07	0.02	0.1	0.02	0.17	0.07	0.15	0.07	0.21	0.17	0.09	0.23	0.02	0.16	0.27
Number Of Fields (NOF)																																		
Percentile Medium Risk																																		
1	1	0.99	0.57	0.4	1	1	1	1	0.76	0.68	0.97	0	1	1	0.75	0.69	0.98	0	1	0.98	0.99	0.95	0	1	1	0.47	0	1	0.43	0	1	0.02	0.99	0.07
-0.02	-0.02	-0.04	-0.12	-0.09	0.04	-0.07	0	-0.02	-0.1	-0.08	0.06	0.34	-0.06	-0.03	-0.1	-0.08	0.05	0.34	-0.06	-0.08	-0.05	0.08	0.37	-0.03	0.03	0.16	0.45	0.04	0.13	0.42	0.02	0.29	-0.11	-0.4
-0.1	-0.1	-0.19	-0.31	-0.23	-0.12	-0.44	-0.08	-0.17	-0.29	-0.21	-0.1	0.13	-0.43	-0.18	-0.3	-0.21	-0.11	0.12	-0.43	-0.31	-0.24	-0.12	0.12	-0.42	-0.19	-0.08	0.17	-0.36	-0.06	0.18	-0.36	0.03	-0.51	-0.82
0.06	0.07	0.11	0.07	0.04	0.2	0.29	0.09	0.12	0.09	0.06	0.21	0.56	0.31	0.13	0.09	0.06	0.22	0.56	0.31	0.15	0.13	0.28	0.62	0.36	0.24	0.39	0.72	0.45	0.32	0.66	0.4	0.55	0.28	0.02

Statistic	Measure	B-A	C-A	D-A	E-A	F-A	G-A	I-A	C-B	D-B	E-B	F-B	G-B	H-B	I-B	D-C	E-C	F-C	G-C	H-C	I-C	E-D	F-D	G-D	H-D	I-D	F-E	G-E	H-E	I-E	G-F	H-F	I-F	H-G	I-G	I-H
Number Of Private Fields (NOPF)																																				
Minimum	pvl	1	1	0	1	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	0.1	0.01	0.04	0.19	0.75	1	1	1	1	1	1	1	1	1	1
	dif	0	0	0.14	0	0	0	0	0	0.14	0	0	0	0	0	0.14	0	0	0	0	0	-0.14	-0.14	-0.14	-0.14	-0.14	0	0	0	0	0	0	0	0	0	0
	lob	-0.05	-0.06	0.04	-0.13	-0.09	-0.11	-0.25	-0.06	0.04	-0.13	-0.09	-0.11	-0.15	-0.25	0.04	-0.13	-0.09	-0.11	-0.15	-0.25	-0.3	-0.27	-0.28	-0.31	-0.41	-0.15	-0.16	-0.19	-0.28	-0.13	-0.17	-0.26	-0.18	-0.27	-0.29
	upb	0.05	0.06	0.24	0.13	0.09	0.11	0.25	0.06	0.24	0.13	0.09	0.11	0.15	0.25	0.25	0.13	0.09	0.11	0.15	0.25	0.01	-0.02	0	0.03	0.12	0.15	0.16	0.19	0.28	0.13	0.17	0.26	0.18	0.27	0.29
Maximum	pvl	0.95	1	0.21	1	0.4	1	1	0.99	0.04	1	0.09	0.93	1	1	0.18	1	0.35	1	1	1	0.34	1	0.92	0.42	0.96	0.52	0.98	1	1	0.99	0.6	0.99	0.98	1	1
	dif	-1.1	-0.2	4.36	-1.68	3.4	1.24	-0.93	0.89	5.45	-0.58	4.5	2.33	-0.83	0.17	4.56	-1.48	3.61	1.44	-1.73	-0.73	-6.04	-0.95	-3.12	-6.29	-5.29	5.08	2.92	-0.25	0.75	-2.17	-5.33	-4.33	-3.17	-2.17	1
	lob	-3.96	-3.25	-1	-8.55	-1.42	-4.49	-14.21	-2.16	0.09	-7.45	-0.32	-3.4	-8.68	-13.12	-0.9	-8.43	-1.33	-4.39	-9.64	-14.05	-14.26	-7.57	-10.42	-15.34	-19.32	-2.81	-5.56	-10.28	-13.93	-9.09	-14.08	-18.17	-12.45	-16.35	-14.16
	upb	1.77	2.85	9.72	5.19	8.23	6.97	12.35	3.95	10.81	6.29	9.32	8.06	7.01	13.45	10.02	5.47	8.54	7.27	6.19	12.6	2.19	5.66	4.18	2.77	8.75	12.97	11.39	9.78	15.43	4.75	3.42	9.5	6.12	12.01	16.16
Median	pvl	1	1	0	1	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	0.04	0	0.04	0.09	0.63	1	1	1	1	1	1	1	1	1	1
	dif	0.02	0.02	1.43	0	0	0.17	0	-0.01	1.4	-0.02	-0.02	0.14	-0.02	-0.02	1.41	-0.02	-0.02	0.15	-0.02	-0.02	-1.43	-1.43	-1.26	-1.43	-1.43	0	0.17	0	0	0.17	0	0	-0.17	-0.17	0
	lob	-0.47	-0.51	0.51	-1.17	-0.82	-0.81	-2.27	-0.53	0.49	-1.2	-0.85	-0.84	-1.36	-2.29	0.48	-1.2	-0.86	-0.84	-1.37	-2.29	-2.83	-2.56	-2.51	-2.97	-3.82	-1.35	-1.28	-1.71	-2.51	-1.01	-1.49	-2.36	-1.75	-2.59	-2.59
	upb	0.51	0.54	2.34	1.17	0.82	1.14	2.27	0.51	2.32	1.15	0.8	1.12	1.32	2.24	2.35	1.17	0.83	1.15	1.34	2.26	-0.02	-0.3	-0.02	0.12	0.97	1.35	1.61	1.71	2.51	1.35	1.49	2.36	1.42	2.25	2.59
Mean	pvl	1	1	0	1	1	1	1	1	0	1	1	0.99	1	1	0	1	1	1	1	1	0.02	0.01	0.07	0.06	0.57	1	1	1	1	1	1	1	1	1	1
	dif	-0.07	0	1.87	-0.12	0.17	0.31	-0.1	0.07	1.94	-0.05	0.24	0.38	-0.06	-0.03	1.87	-0.13	0.17	0.3	-0.14	-0.1	-2	-1.7	-1.57	-2.01	-1.97	0.3	0.43	-0.01	0.03	0.13	-0.31	-0.27	-0.44	-0.4	0.04
	lob	-0.72	-0.69	0.67	-1.67	-0.91	-0.98	-3.09	-0.62	0.74	-1.6	-0.84	-0.91	-1.83	-3.02	0.64	-1.69	-0.94	-1.01	-1.92	-3.1	-3.85	-3.19	-3.21	-4.05	-5.13	-1.48	-1.48	-2.27	-3.28	-1.43	-2.28	-3.39	-2.53	-3.6	-3.38
	upb	0.58	0.69	3.08	1.42	1.26	1.6	2.9	0.76	3.15	1.49	1.33	1.67	1.7	2.97	3.1	1.44	1.28	1.62	1.65	2.9	-0.15	-0.21	0.08	0.03	1.19	2.07	2.34	2.25	3.33	1.69	1.66	2.85	1.65	2.79	3.45
Variance	pvl	1	1	0	1	0.96	1	1	1	0	1	0.92	1	1	1	0	1	0.94	1	1	1	0.11	0.23	0.09	0.2	0.77	0.99	1	1	1	1	1	1	1	1	1
	dif	-0.51	-0.29	16.74	-0.7	4.31	0.82	-0.67	0.22	17.25	-0.19	4.82	1.33	-0.2	-0.16	17.03	-0.41	4.6	1.1	-0.42	-0.38	-17.44	-12.43	-15.93	-17.45	-17.41	5.01	1.51	-0.01	0.03	-3.5	-5.02	-4.98	-1.52	-1.49	0.04
	lob	-7.25	-7.47	4.14	-16.85	-7.02	-12.66	-31.9	-6.96	4.65	-16.34	-6.51	-12.15	-18.64	-31.39	4.18	-16.75	-7.01	-12.6	-19.04	-31.72	-36.79	-27.98	-33.1	-38.75	-50.41	-13.54	-18.41	-23.59	-34.49	-19.77	-25.6	-37.52	-23.35	-34.83	-35.61
	upb	6.23	6.9	29.34	15.46	15.65	14.29	30.56	7.41	29.85	15.97	16.16	14.8	18.25	31.07	29.87	15.93	16.21	14.8	18.19	30.95	1.91	3.13	1.25	3.85	15.59	23.56	21.44	23.57	34.54	12.77	15.56	27.55	20.3	31.86	35.68
Standard Deviation	pvl	0.98	1	0.01	1	0.77	1	1	1	0	1	0.36	0.95	1	1	0	1	0.69	1	1	1	0.07	0.63	0.39	0.12	0.77	0.81	0.98	1	1	1	0.85	1	0.99	1	1
	dif	-0.25	-0.07	1.72	-0.38	0.67	0.34	-0.24	0.18	1.97	-0.13	0.93	0.6	-0.18	0.01	1.79	-0.31	0.74	0.41	-0.36	-0.17	-2.1	-1.04	-1.37	-2.15	-1.96	1.05	0.73	-0.05	0.14	-0.33	-1.11	-0.91	-0.78	-0.58	0.19
	lob	-1.01	-0.88	0.3	-2.2	-0.6	-1.17	-3.76	-0.62	0.55	-1.95	-0.35	-0.92	-2.26	-3.5	0.34	-2.15	-0.56	-1.13	-2.46	-3.7	-4.28	-2.8	-3.31	-4.55	-5.67	-1.03	-1.52	-2.71	-3.74	-2.16	-3.42	-4.58	-3.23	-4.34	-3.82
	upb	0.5	0.74	3.14	1.44	1.95	1.86	3.28	0.99	3.39	1.69	2.2	2.11	1.9	3.53	3.23	1.53	2.05	1.95	1.73	3.36	0.08	0.71	0.56	0.25	1.76	3.14	2.97	2.6	4.03	1.5	1.21	2.75	1.68	3.17	4.2
Alves Low Risk	pvl	1	1	0	1	1	1	1	1	0.02	1	1	1	1	1	0.01	1	1	1	1	1	0.17	0.05	0.3	0.12	0.57	1	1	1	1	1	1	1	0.99	1	1
	dif	-0.02	-0.01	-0.23	0.01	0	-0.04	0.07	0.01	-0.21	0.03	0.02	-0.02	0.07	0.1	-0.22	0.02	0.01	-0.03	0.06	0.08	0.24	0.23	0.19	0.28	0.3	-0.01	-0.05	0.04	0.06	-0.04	0.05	0.07	0.09	0.11	0.02
	lob	-0.12	-0.12	-0.41	-0.23	-0.16	-0.24	-0.39	-0.09	-0.39	-0.2	-0.14	-0.21	-0.2	-0.36	-0.41	-0.22	-0.16	-0.23	-0.21	-0.38	-0.04	0	-0.06	-0.03	-0.18	-0.28	-0.34	-0.31	-0.45	-0.28	-0.26	-0.41	-0.23	-0.38	-0.5
	upb	0.08	0.09	-0.04	0.25	0.17	0.16	0.53	0.12	-0.02	0.27	0.19	0.18	0.34	0.55	-0.03	0.26	0.18	0.17	0.33	0.54	0.53	0.46	0.44	0.59	0.79	0.26	0.24	0.38	0.57	0.2	0.35	0.55	0.41	0.6	0.55
Alves Very High Risk	pvl	1	1	0	1	1	0.84	1	1	0	1	1	0.69	1	1	0	1	1	0.85	1	1	0	0	0.11	0.01	0.26	1	0.87	1	1	0.98	1	1	0.87	0.99	1
	dif	-0.01	0	0.25	-0.03	0.01	0.08	-0.04	0.01	0.27	-0.02	0.02	0.09	-0.03	-0.03	0.25	-0.03	0.01	0.08	-0.04	-0.04	-0.29	-0.24	-0.18	-0.3	-0.3	0.04	0.11	-0.01	-0.01	0.06	-0.05	-0.05	-0.12	-0.12	0
	lob	-0.09	-0.08	0.11	-0.22	-0.12	-0.08	-0.4	-0.07	0.12	-0.2	-0.11	-0.07	-0.24	-0.39	0.11	-0.22	-0.12	-0.08	-0.26	-0.4	-0.51	-0.42	-0.38	-0.54	-0.68	-0.17	-0.12	-0.28	-0.41	-0.12	-0.29	-0.43	-0.37	-0.5	-0.41
	upb	0.06	0.08	0.4	0.15	0.14	0.23	0.32	0.09	0.41	0.17	0.15	0.24	0.18	0.33	0.4	0.16	0.15	0.23	0.17	0.32	-0.06	-0.06	0.02	-0.05	0.08	0.26	0.34	0.26	0.39	0.25	0.18	0.32	0.13	0.27	0.41
Number Of Methods (NOM)																																				
Percentile High Risk	pvl	0.75	1	0.57	1	1	1	1	0.51	0.11	1	1	1	0.03	1	0.77	0.99	0.99	1	0.26	1	0.59	0.47	0.72	0.97	0.97	1	1	0.18	1	1	0.13	1	0.24	1	0.74
	dif	-0.03	0.01	0.07	-0.04	-0.02	-0.02	-0.04	0.04	0.11	0	0.01	0.02	0.18	0	0.06	-0.05	-0.03	-0.03	0.14	-0.05	-0.11	-0.1	-0.09	0.07	-0.11	0	0.02	0.19	0	0.01	0.17	-0.01	0.16	-0.02	-0.18
	lob	-0.1	-0.06	-0.05	-0.19	-0.13	-0.14	-0.33	-0.02	-0.01	-0.16	-0.1	-0.11	0.01	-0.3	-0.06	-0.2	-0.15	-0.16	-0.04	-0.34	-0.3	-0.25	-0.25	-0.13	-0.42	-0.16	-0.17	-0.04	-0.33	-0.15	-0.02	-0.32	-0.04	-0.34	-0.52
	upb	0.03	0.08	0.19	0.11	0.08	0.11	0.26	0.11	0.23	0.15	0.12	0.15	0.36	0.3	0.19	0.11	0.08	0.1	0.32	0.25	0.07	0.05	0.07	0.28	0.2	0.19	0.21	0.41	0.33	0.16	0.37	0.3	0.37	0.3	0.16
Alves Very High Risk	pvl	0.73	0.95	0.99	1	0.03	1	1	1	0.65	0.88	0.34	0.98	0.94	1	0.81	0.95	0.23	1	0.97	1	1	0.05	1	1	1	0.17	1	1	1	0.27	0.29	0.84	1	1	1
	dif	0.01	0.01	-0.01	-0.01	0.04	0	-0.01	0	-0.02	-0.02	0.02	-0.01	-0.02	-0.02	-0.02	-0.02	0.03	-0.01	-0.02	-0.02	0	0.05	0.01	0	0	0.05	0.01	0	0	-0.04	-0.05	-0.05	-0.01	-0.01	0
	lob	-0.01	-0.01	-0.05	-0.06	0	-0.04	-0.1	-0.02	-0.06	-0.07	-0.01	-0.05	-0.08	-0.11	-0.06	-0.07	-0.01	-0.05	-0.07	-0.11	-0.06	0	-0.04	-0.06	-0.1	-0.01	-0.05	-0.07	-0.1	-0.09	-0.11	-0.14	-0.07	-0.11	-0.11
	upb	0.03	0.03	0.03	0.04	0.07	0.04	0.08	0.02	0.02	0.03	0.06	0.03	0.03	0.07	0.02	0.03	0.06	0.03	0.04	0.07	0.06	0.09	0.06	0.06	0.1	0.1	0.07	0.07	0.1	0.01	0.01	0.05	0.06	0.09	0.11
Number Of Private Methods (NOPM)																																				
Minimum	pvl	0.98	1	0.4	0.61	0.45	0.5	0.99	1	0.13	0.87	0.84	0.2	1	0.97	0.18	0.84	0.81	0.26	1	0.98	0.07	0.03	1	0.99	1	1	0.09	0.77	0.75	0.04	0.81	0.8	0.99	1	1
	dif	-0.12	-0.1	0.48	-0.52	-0.41	0.48	0.48	0.02	0.6	-0.4	-0.29	0.6	0.26	0.6	0.58	-0.42	-0.31	0.58	0.24	0.58	-1	-0.89	0	-0.33	0	0.11	1	0.67	1	0.89	0.56	0.89	-0.33	0	0.33
	lob	-0.48	-0.49	-0.2	-1.39	-1.02	-0.25	-1.2	-0.37	-0.08	-1.27	-0.9	-0.13	-0.73	-1.08	-0.11	-1.3	-0.94	-0.16	-0.76	-1.11	-2.04	-1.72	-0.92	-1.48	-1.77	-0.89	-0.07	-0.6	-0.85	0.01	-0.55	-0.86	-1.51	-1.79	-1.58
	upb	0.24	0.29	1.15	0.34	0.2	1.2	2.15	0.41	1.27	0.46	0.32	1.32	1.25	2.27	1.27	0.45	0.31	1.31	1.24	2.26	0.04	-0.05	0.92	0.81	1.77	1.11	2.07	1.93	2.85	1.76	1.66	2.64	0.84	1.79	2.25
Percentile High Risk	pvl	0.95	0.94	0.97	1	1	1	1	0.29	0.71	1	0.98	1	0	1	1	1	1	0.95	0	1	0.99	1	0.94	0.01	1	1	1	0	1	1	0	1	0	1	0.17
	dif	-0.03	0.03	0.05	-0.01	0.01	-0.03	-0.01	0.07	0.08	0.02	0.04	0	0.38	0.02	0.02	-0.05	-0.02	-0.06	0.31	-0.05	-0.07	-0.04	-0.08	0.29	-0.06	0.03	-0.02	0.36	0	-0.04	0.33	-0.02	0.37	0.02	-0.36
	lob	-0.11	-0.05	-0.1	-0.2	-0.12	-0.19	-0.38	-0.02	-0.07	-0.17	-0.09	-0.16	0.16	-0.35	-0.13	-0.24	-0.16	-0.23	0.09	-0.42	-0.29	-0.22	-0.29	0.04	-0.46	-0.19	-0.25	0.08	-0.41	-0.23	0.09	-0.41	0.12	-0.38	-0.78
	upb	0.05	0.12	0.2	0.18	0.15	0.13	0.36	0.15	0.23	0.21	0.18	0.16	0.59	0.39	0.17	0.15	0.12	0.1	0.53	0.32	0.16	0.14	0.12	0.54	0.33	0.24	0.22	0.64	0.41	0.15	0.58	0.36	0.63	0.41	0.07
Alves Low Risk	pvl	1	1	0.11	1	0.21	1	0.99	0.91	0.27	1	0.48	1	0.8	0.98	0.05	1	0.1	1	0.98	1	0.47	1	0.28	0.12	0.61	0.67	1	1	1	0.46	0.2	0.73	0.99	1	1
	dif	-0.01	0.01	-0.08	0.01	-0.06	0.01	0.06	0.02	-0.07	0.02	-0.05	0.02	0.06	0.07	-0.09	0	-0.07	0	0.04	0.05	0.09	0.02	0.09	0.13	0.14	-0.07	0	0.04	0.05	0.08	0.12	0.12	0.04	0.05	0.01
	lob	-0.06	-0.04	-0.17	-0.1	-0.14	-0.08	-0.16	-0.03	-0.16	-0.09	-0.13	-0.07	-0.06	-0.15	-0.18	-0.11	-0.15	-0.09	-0.09	-0.17	-0.04	-0.09	-0.03	-0.02	-0.09	-0.2	-0.14	-0.12	-0.19	-0.04	-0.03	-0.1	-0.11	-0.18	-0.24
	upb	0.03	0.06	0.01	0.12	0.01	0.11	0.28	0.07	0.02	0.13	0.03	0.12	0.19	0.29	0	0.11	0.01	0.1	0.17	0.27	0.22	0.12	0.21	0.28	0.37	0.05	0.14	0.21	0.29	0.19	0.26	0.35	0.19	0.28	0.25
Alves Medium Risk	pvl	0.99	1	0.05	1	0.52	1	1	0.97	0.17	1	0.87	1	1	0.98	0.04	1	0.46	1	1	1	0.45	0.98	0.29	0.65	0.54	0.92	1	1	1	0.85	0.97	0.84	1	1	1
	dif	0.01	0	0.07	0	0.04	0	-0.04	-0.01	0.06	-0.01	0.03	-0.01	-0.01	-0.05	0.07	0	0.04	0	0	-0.04	-0.07	-0.03	-0.07	-0.07	-0.11	0.04	0	0	-0.04	-0.04	-0.04	-0.08	0	-0.04	-0.04
	lob	-0.02	-0.04	0	-0.09	-0.02	-0.07	-0.21	-0.05	-0.01	-0.1	-0.03	-0.08	-0.11	-0.22	0	-0.09	-0.02	-0.07	-0.09	-0.2	-0.17	-0.11	-0.16	-0.18	-0.28	-0.06	-0.1	-0.12	-0.22	-0.13	-0.15	-0.25	-0.11	-0.22	-0.23
	upb	0.05	0.03	0.13	0.08	0.1	0.07	0.12	0.02	0.12	0.07	0.09	0.06	0.09	0.11	0.14	0.09	0.1	0.07	0.1	0.12	0.03	0.05	0.02	0.05	0.06	0.14	0.1	0.13	0.14	0.04	0.07	0.09	0.12	0.13	0.14

B-A	C-A	D-A	E-A	F-A	G-A	I-A	C-B	D-B	E-B	F-B	G-B	H-B	I-B	D-C	E-C	F-C	G-C	H-C	I-C	E-D	F-D	G-D	H-D	I-D	F-E	G-E	H-E	I-E	G-F	H-F	I-F	H-G	I-G	I-H
pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb	pvl dif lob upb
Type Lines of Code (TLOC)																																		
Minimum																																		
0.99	0.95	0.95	1	1	0.06	1	1	1	1	1	0.01	0.14	1	1	0.99	1	0.01	0.1	1	0.98	1	0.03	0.12	1	1	0.67	0.77	1	0.14	0.35	1	1	0.97	0.96
-0.79	-1.09	-1.86	0.68	-0.35	5.1	0.43	-0.3	-1.07	1.46	0.44	5.88	6.21	1.21	-0.77	1.77	0.74	6.18	6.52	1.52	2.54	1.51	6.95	7.29	2.29	-1.03	4.42	4.75	-0.25	5.44	5.78	0.78	0.33	-4.67	-5
-3.37	-3.84	-6.69	-5.52	-4.7	-0.08	-11.56	-3.06	-5.91	-4.74	-3.92	0.71	-0.87	-10.78	-5.7	-4.51	-3.72	0.92	-0.63	-10.51	-4.89	-4.46	0.36	-0.89	-10.38	-8.15	-3.23	-4.3	-13.5	-0.8	-2.12	-11.71	-8.05	-17.47	-18.68
1.8	1.67	2.98	6.88	4	10.27	12.42	2.46	3.77	7.66	4.79	11.05	13.3	13.2	4.16	8.04	5.19	11.44	13.66	13.54	9.96	7.48	13.54	15.46	14.95	6.09	12.07	13.8	13	11.69	13.68	13.27	8.71	8.13	8.68
Percentile Medium Risk																																		
0.77	0.99	0.74	1	1	0.56	0.83	0.26	0.2	1	1	0.12	1	0.61	0.96	0.98	0.98	0.87	1	0.92	0.78	0.71	1	0.89	1	1	0.65	1	0.75	0.55	1	0.76	0.79	1	0.8
-0.04	0.02	0.07	-0.03	-0.02	0.09	0.16	0.06	0.11	0	0.02	0.12	0.01	0.19	0.05	-0.05	-0.04	0.07	-0.05	0.14	-0.1	-0.09	0.02	-0.1	0.09	0.01	0.12	0	0.19	0.11	-0.01	0.18	-0.12	0.07	0.19
-0.11	-0.05	-0.06	-0.2	-0.14	-0.05	-0.16	-0.02	-0.02	-0.16	-0.1	-0.02	-0.18	-0.13	-0.08	-0.22	-0.16	-0.07	-0.24	-0.18	-0.3	-0.25	-0.16	-0.32	-0.25	-0.18	-0.08	-0.24	-0.16	-0.06	-0.22	-0.16	-0.34	-0.27	-0.18
0.03	0.09	0.2	0.13	0.1	0.22	0.48	0.13	0.23	0.17	0.13	0.26	0.2	0.51	0.18	0.11	0.08	0.21	0.14	0.46	0.1	0.07	0.19	0.12	0.43	0.2	0.32	0.25	0.54	0.27	0.2	0.51	0.11	0.41	0.55
Percentile High Risk																																		
1	0.61	0.84	1	0.97	0.96	0.98	0.63	0.85	1	0.97	0.96	0.45	0.98	1	0.93	0.36	0.45	0.11	1	0.9	0.5	0.53	0.14	1	1	1	0.92	0.96	1	0.94	0.9	0.98	0.88	0.54
0	-0.03	-0.04	0.02	0.03	0.03	-0.07	-0.03	-0.04	0.02	0.03	0.03	0.09	-0.07	-0.01	0.05	0.06	0.06	0.12	-0.04	0.06	0.07	0.08	0.13	-0.03	0.01	0.02	0.07	-0.09	0.01	0.06	-0.1	0.05	-0.1	-0.16
-0.05	-0.08	-0.13	-0.09	-0.05	-0.06	-0.29	-0.08	-0.13	-0.09	-0.05	-0.06	-0.04	-0.28	-0.1	-0.07	-0.02	-0.03	-0.01	-0.26	-0.07	-0.04	-0.04	-0.02	-0.26	-0.12	-0.12	-0.09	-0.32	-0.11	-0.08	-0.32	-0.1	-0.33	-0.4
0.05	0.02	0.04	0.13	0.11	0.13	0.14	0.02	0.05	0.13	0.11	0.13	0.21	0.14	0.08	0.16	0.14	0.16	0.24	0.17	0.19	0.18	0.19	0.27	0.2	0.14	0.15	0.23	0.15	0.12	0.2	0.13	0.2	0.12	0.09
Number of Children (NC)																																		
Alves Medium Risk																																		
0.84	1	0.02	0.98	0.93	1	1	0.71	0.17	0.78	1	0.94	0.91	1	0.01	0.99	0.86	1	0.53	1	0.06	0.63	0.08	1	0.98	0.76	1	0.42	1	0.92	0.99	1	0.58	1	1
0.02	0	0.1	-0.04	0.03	-0.02	0.02	-0.03	0.07	-0.06	0.01	-0.04	0.05	0	0.1	-0.03	0.04	-0.01	0.08	0.03	-0.13	-0.06	-0.11	-0.02	-0.08	0.07	0.02	0.11	0.06	-0.05	0.04	-0.01	0.09	0.04	-0.06
-0.02	-0.05	0.01	-0.15	-0.05	-0.11	-0.19	-0.08	-0.01	-0.17	-0.07	-0.13	-0.07	-0.22	0.01	-0.14	-0.04	-0.1	-0.05	-0.19	-0.26	-0.17	-0.23	-0.16	-0.3	-0.06	-0.12	-0.05	-0.18	-0.16	-0.1	-0.23	-0.06	-0.19	-0.3
0.07	0.04	0.18	0.08	0.11	0.08	0.23	0.02	0.16	0.05	0.09	0.05	0.18	0.21	0.19	0.08	0.12	0.08	0.21	0.24	0	0.04	0.01	0.13	0.15	0.19	0.16	0.27	0.29	0.06	0.19	0.21	0.24	0.26	0.19
Depth of Inheritance Tree (DIT)																																		
Percentile Low Risk																																		
1	1	0.03	0.99	1	0.83	1	1	0.02	1	1	0.89	0.98	1	0.01	1	1	0.95	0.99	1	0.1	0.28	0.01	0.09	0.62	0.99	1	1	1	0.84	0.95	1	1	1	1
-0.02	-0.05	0.47	-0.16	0.04	-0.24	-0.24	-0.03	0.49	-0.14	0.06	-0.22	-0.22	-0.22	0.52	-0.11	0.09	-0.19	-0.19	-0.19	-0.63	-0.43	-0.71	-0.71	-0.71	0.2	-0.08	-0.08	-0.08	-0.28	-0.28	-0.28	0	0	0
-0.26	-0.31	0.02	-0.74	-0.36	-0.72	-1.35	-0.29	0.04	-0.72	-0.34	-0.7	-0.87	-1.33	0.06	-0.69	-0.32	-0.68	-0.85	-1.31	-1.32	-0.98	-1.32	-1.47	-1.89	-0.46	-0.79	-0.92	-1.31	-0.86	-1.01	-1.44	-0.78	-1.19	-1.27
0.22	0.21	0.92	0.42	0.45	0.24	0.88	0.23	0.94	0.44	0.47	0.26	0.44	0.9	0.98	0.47	0.51	0.3	0.48	0.93	0.06	0.13	-0.1	0.05	0.47	0.87	0.63	0.76	1.15	0.3	0.45	0.88	0.78	1.19	1.27
Percentile Very High Risk																																		
1	1	0.04	1	1	0.79	1	1	0.03	1	1	0.86	0.97	1	0.04	1	1	0.85	0.97	1	0.19	0.31	0.01	0.09	0.62	1	1	1	1	0.82	0.94	1	1	1	1
0.02	0.01	-0.53	0.13	-0.05	0.28	0.28	-0.01	-0.55	0.11	-0.07	0.26	0.26	0.26	-0.54	0.12	-0.06	0.27	0.27	0.27	0.66	0.48	0.81	0.81	0.81	-0.18	0.15	0.15	0.15	0.33	0.33	0.33	0	0	0
-0.25	-0.28	-1.05	-0.53	-0.51	-0.27	-1	-0.3	-1.07	-0.55	-0.54	-0.29	-0.49	-1.02	-1.07	-0.55	-0.54	-0.29	-0.49	-1.01	-0.13	-0.16	0.11	-0.06	-0.54	-0.94	-0.66	-0.81	-1.26	-0.33	-0.51	-1	-0.89	-1.37	-1.46
0.3	0.31	-0.01	0.79	0.42	0.84	1.56	0.29	-0.03	0.77	0.39	0.81	1.02	1.54	-0.02	0.79	0.41	0.83	1.03	1.55	1.45	1.12	1.52	1.69	2.17	0.58	0.97	1.12	1.57	1	1.18	1.67	0.89	1.37	1.46
Alves Low Risk																																		
0.88	0.42	1	0.23	1	1	1	1	0.9	0.6	1	0.97	0.53	1	0.66	0.84	1	0.85	0.33	1	0.24	0.99	1	0.99	1	0.63	0.35	0.1	0.97	1	0.78	1	0.98	1	1
-0.06	-0.09	0.05	-0.25	-0.04	0.03	-0.01	-0.04	0.11	-0.19	0.02	0.09	0.23	0.05	0.14	-0.15	0.06	0.12	0.26	0.08	-0.29	-0.08	-0.02	0.12	-0.06	0.21	0.28	0.41	0.23	0.07	0.2	0.02	0.14	-0.04	-0.18
-0.19	-0.23	-0.19	-0.55	-0.25	-0.23	-0.61	-0.17	-0.13	-0.49	-0.19	-0.17	-0.13	-0.55	-0.1	-0.46	-0.16	-0.14	-0.09	-0.51	-0.66	-0.38	-0.34	-0.29	-0.69	-0.14	-0.1	-0.04	-0.42	-0.24	-0.19	-0.6	-0.28	-0.68	-0.86
0.07	0.04	0.29	0.06	0.18	0.29	0.58	0.1	0.35	0.12	0.24	0.34	0.58	0.64	0.39	0.16	0.28	0.39	0.62	0.68	0.07	0.21	0.31	0.52	0.57	0.56	0.65	0.86	0.89	0.38	0.59	0.64	0.55	0.59	0.5
Alves Medium Risk																																		
0.57	0.1	0.44	0.31	0.99	0.74	1	0.98	0.97	0.83	1	1	0.43	0.98	1	0.98	0.99	1	0.22	0.94	1	0.97	1	0.22	0.9	0.84	1	0.14	0.78	1	0.69	0.99	0.36	0.94	1
0.03	0.04	0.05	0.07	0.02	0.05	-0.04	0.01	0.03	0.05	-0.01	0.02	-0.08	-0.06	0.01	0.03	-0.02	0	-0.09	-0.08	0.02	-0.03	-0.01	-0.11	-0.09	-0.06	-0.03	-0.13	-0.11	0.03	-0.07	-0.06	-0.1	-0.08	0.02
-0.02	0	-0.02	-0.02	-0.05	-0.04	-0.23	-0.03	-0.05	-0.05	-0.08	-0.06	-0.19	-0.26	-0.07	-0.07	-0.09	-0.08	-0.21	-0.27	-0.1	-0.13	-0.11	-0.24	-0.29	-0.17	-0.15	-0.27	-0.32	-0.07	-0.2	-0.26	-0.23	-0.29	-0.2
0.07	0.08	0.13	0.17	0.09	0.13	0.16	0.06	0.11	0.15	0.06	0.1	0.04	0.13	0.09	0.14	0.05	0.09	0.02	0.12	0.14	0.06	0.1	0.03	0.11	0.06	0.09	0.02	0.1	0.13	0.05	0.15	0.04	0.12	0.24
Lack of Cohesion Of Methods (LCOM)																																		
Percentile Medium Risk																																		
1	1	0.09	1	0.75	0.98	1	0.94	0.04	1	0.54	1	1	1	0.26	1	0.96	0.89	0.98	1	0.41	0.96	0.08	0.3	0.83	0.92	1	1	1	0.52	0.8	0.98	1	1	1
-0.01	0.01	0.09	-0.01	0.04	-0.03	-0.03	0.02	0.1	0	0.05	-0.02	-0.02	-0.02	0.07	-0.03	0.03	-0.05	-0.05	-0.05	-0.1	-0.04	-0.12	-0.12	-0.12	0.06	-0.02	-0.02	-0.02	-0.08	-0.08	-0.08	0	0	0
-0.06	-0.04	-0.01	-0.13	-0.04	-0.13	-0.26	-0.03	0	-0.12	-0.03	-0.12	-0.16	-0.25	-0.02	-0.15	-0.05	-0.15	-0.18	-0.28	-0.24	-0.16	-0.25	-0.28	-0.36	-0.08	-0.17	-0.15	-0.27	-0.2	-0.23	-0.32	-0.16	-0.25	-0.26
0.04	0.07	0.18	0.11	0.13	0.07	0.2	0.07	0.19	0.11	0.14	0.08	0.11	0.21	0.17	0.09	0.12	0.06	0.09	0.19	0.04	0.07	0.01	0.04	0.12	0.19	0.13	0.15	0.24	0.04	0.07	0.16	0.16	0.25	0.26
Alves Medium Risk																																		
1	1	0.23	1	0.09	0.74	1	1	0.5	1	0.27	0.94	0.75	1	0.37	1	0.18	0.86	0.85	1	0.96	1	1	0.16	0.99	0.93	1	0.87	1	1	0.11	0.98	0.4	1	1
0.01	0	0.05	0.01	0.05	0.03	0	0	0.04	0.01	0.04	0.03	-0.05	-0.01	0.04	0.01	0.05	0.03	-0.04	0	-0.03	0	-0.01	-0.09	-0.05	0.04	0.02	-0.05	-0.01	-0.02	-0.09	-0.05	-0.07	-0.03	0.04
-0.02	-0.03	-0.01	-0.06	0	-0.03	-0.15	-0.04	-0.02	-0.07	-0.01	-0.04	-0.13	-0.15	-0.02	-0.07	-0.01	-0.03	-0.13	-0.15	-0.12	-0.07	-0.09	-0.19	-0.2	-0.05	-0.07	-0.16	-0.18	-0.09	-0.18	-0.2	-0.17	-0.19	-0.13
0.04	0.04	0.11	0.09	0.1	0.1	0.15	0.03	0.1	0.08	0.09	0.09	0.04	0.14	0.1	0.09	0.1	0.09	0.05	0.14	0.06	0.08	0.07	0.01	0.11	0.12	0.11	0.06	0.15	0.06	0.01	0.1	0.03	0.12	0.21
Alves High Risk																																		
0.9	1	0.14	0.93	1	0.99	0.99	0.53	0.02	1	0.77	1	1	1	0.34	0.81	1	0.92	0.9	0.97	0.1	0.79	0.13	0.19	0.58	0.78	1	1	1	0.9	0.87	0.96	1	1	1
-0.02	0.01	0.09	-0.05	0.02	-0.03	-0.08	0.04	0.11	-0.03	0.05	-0.01	-0.03	-0.05	0.08	-0.07	0.01	-0.05	-0.07	-0.09	-0.14	-0.06	-0.12	-0.14	-0.16	0.08	0.02	0	-0.02	-0.06	-0.08	-0.1	-0.02	-0.04	-0.02
-0.08	-0.04	-0.01	-0.18	-0.07	-0.14	-0.33	-0.02	0.01	-0.16	-0.04	-0.12	-0.18	-0.3	-0.03	-0.2	-0.08	-0.16	-0.22	-0.34	-0.3	-0.19	-0.26	-0.31	-0.43	-0.07	-0.14	-0.19	-0.3	-0.19	-0.24	-0.36	-0.2	-0.31	-0.31
0.03	0.07	0.19	0.08	0.12	0.07	0.18	0.09	0.21	0.1	0.14	0.1	0.12	0.2	0.18	0.07	0.1	0.06	0.08	0.16	0.01	0.06	0.02	0.03	0.1	0.23	0.18	0.19	0.26	0.07	0.09	0.16	0.16	0.23	0.27

	B-A pvl dif lob upb	C-A pvl dif lob upb	D-A pvl dif lob upb	E-A pvl dif lob upb	F-A pvl dif lob upb	G-A pvl dif lob upb	I-A pvl dif lob upb	C-B pvl dif lob upb	D-B pvl dif lob upb	E-B pvl dif lob upb	F-B pvl dif lob upb	G-B pvl dif lob upb	H-B pvl dif lob upb	I-B pvl dif lob upb	D-C pvl dif lob upb	E-C pvl dif lob upb	F-C pvl dif lob upb	G-C pvl dif lob upb
Fan-in (FANIN)																		
Minimum																		
	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1
	0	0	0	0	0	0	0	0	0	0	0	0	0.33	0	0	0	0	0
	-0.05	-0.05	-0.09	-0.11	-0.08	-0.1	-0.22	-0.05	-0.09	-0.11	-0.08	-0.1	0.2	-0.22	-0.09	-0.12	-0.08	-0.1
	0.05	0.05	0.09	0.11	0.08	0.1	0.22	0.05	0.09	0.11	0.08	0.1	0.46	0.22	0.09	0.12	0.08	0.1
Maximum																		
	0.92	0.68	0.76	1	1	0.76	0.99	1	0.3	1	0.93	0.98	0.55	1	0.17	1	0.79	1
	2.33	3.35	-5.45	1.08	-1.5	5.83	7.83	1.02	-7.79	-1.25	-3.83	3.5	-9.5	5.5	-8.8	-2.27	-4.85	2.48
	-3.14	-2.49	-15.7	-12.05	-10.72	-5.12	-17.56	-4.82	-18.03	-14.38	-13.05	-7.45	-24.5	-19.89	-19.24	-15.55	-14.29	-8.65
	7.81	9.19	4.79	14.22	7.72	16.79	33.23	6.85	2.46	11.88	5.39	14.45	5.5	30.89	1.64	11.02	4.59	13.62
Standard Deviation																		
	0.67	0.61	0.96	1	1	0.7	0.99	1	0.45	1	0.83	0.99	0.36	1	0.4	1	0.79	1
	0.48	0.53	-0.57	0.24	-0.2	0.93	1.21	0.06	-1.05	-0.24	-0.68	0.45	-1.63	0.73	-1.1	-0.29	-0.73	0.4
	-0.35	-0.34	-2.11	-1.73	-1.59	-0.72	-2.6	-0.82	-2.59	-2.21	-2.06	-1.19	-3.88	-3.08	-2.67	-2.29	-2.15	-1.28
	1.3	1.41	0.97	2.21	1.18	2.57	5.02	0.93	0.49	1.74	0.71	2.1	0.62	4.55	0.47	1.7	0.68	2.07
Percentile High Risk																		
	0.53	0.98	0.93	1	1	0.46	0.99	0.99	0.31	1	0.65	0.97	0.02	1	0.67	1	0.94	0.83
	-0.05	-0.03	0.06	-0.02	0.03	-0.11	-0.1	0.02	0.11	0.03	0.08	-0.06	0.24	-0.05	0.09	0.01	0.05	-0.08
	-0.13	-0.11	-0.09	-0.21	-0.1	-0.26	-0.47	-0.06	-0.04	-0.16	-0.05	-0.21	0.02	-0.42	-0.06	-0.18	-0.08	-0.24
	0.03	0.06	0.21	0.17	0.16	0.05	0.26	0.11	0.26	0.22	0.21	0.1	0.45	0.31	0.24	0.2	0.19	0.08
Alves Low Risk																		
	1	0.99	0.24	1	0.97	1	0.98	0.87	0.11	1	0.84	1	0.64	0.99	0.56	0.96	1	1
	0.01	-0.02	-0.09	0.04	-0.04	0.01	0.1	-0.03	-0.11	0.03	-0.05	0	0.1	0.09	-0.07	0.06	-0.02	0.03
	-0.05	-0.09	-0.21	-0.11	-0.14	-0.12	-0.19	-0.1	-0.22	-0.13	-0.16	-0.13	-0.07	-0.21	-0.19	-0.1	-0.13	-0.1
	0.08	0.05	0.02	0.19	0.07	0.13	0.39	0.04	0.01	0.18	0.05	0.12	0.27	0.38	0.04	0.21	0.09	0.15
Alves High Risk																		
	1	1	0	0.99	1	0.97	0.99	0.99	0	1	1	0.99	1	0.99	0.01	0.97	1	0.9
	0	0.01	0.08	-0.02	0	-0.02	-0.05	0.01	0.09	-0.02	0.01	-0.02	-0.02	-0.04	0.08	-0.03	0	-0.03
	-0.04	-0.03	0.02	-0.1	-0.05	-0.09	-0.2	-0.03	0.02	-0.1	-0.05	-0.09	-0.11	-0.2	0.01	-0.11	-0.06	-0.1
	0.03	0.04	0.14	0.06	0.06	0.04	0.11	0.05	0.15	0.06	0.06	0.05	0.07	0.11	0.14	0.05	0.05	0.04
Alves Very High Risk																		
	0.84	1	0.51	0.2	0.66	1	1	1	0.11	0.58	0.99	0.88	0.57	1	0.29	0.38	0.91	0.98
	0.01	0	-0.02	0.04	0.02	-0.01	0.02	0	-0.03	0.03	0.01	-0.02	-0.03	0.01	-0.03	0.03	0.01	-0.01
	-0.01	-0.01	-0.06	-0.01	-0.01	-0.04	-0.07	-0.02	-0.06	-0.02	-0.02	-0.05	-0.08	-0.08	-0.06	-0.01	-0.02	-0.05
	0.03	0.02	0.01	0.08	0.05	0.03	0.1	0.01	0	0.07	0.04	0.02	0.02	0.09	0.01	0.07	0.04	0.02
Fan-out (FANOUT)																		
Alves High Risk																		
	1	1	1	1	0.82	0.01	1	1	1	1	0.88	0.01	0.89	1	1	1	0.92	0.01
	0	0	0	-0.01	0.03	0.09	-0.04	0	0	-0.01	0.03	0.09	0.05	-0.04	0	-0.01	0.03	0.08
	-0.03	-0.04	-0.07	-0.1	-0.03	0.02	-0.2	-0.04	-0.07	-0.1	-0.03	0.01	-0.05	-0.21	-0.07	-0.1	-0.04	0.01
	0.04	0.04	0.07	0.08	0.09	0.16	0.13	0.04	0.07	0.08	0.09	0.16	0.14	0.13	0.07	0.08	0.09	0.16

	H-C pvl dif lob upb	I-C pvl dif lob upb	E-D pvl dif lob upb	F-D pvl dif lob upb	G-D pvl dif lob upb	H-D pvl dif lob upb	I-D pvl dif lob upb	F-E pvl dif lob upb	G-E pvl dif lob upb	H-E pvl dif lob upb	I-E pvl dif lob upb	G-F pvl dif lob upb	H-F pvl dif lob upb	I-F pvl dif lob upb	H-G pvl dif lob upb	I-G pvl dif lob upb	I-H pvl dif lob upb
Fan-in (FANIN)																	
Minimum																	
	0	1	1	1	1	0	1	1	1	0	1	1	0	1	0	1	0
	0.33	0	0	0	0	0.33	0	0	0	0.33	0	0	0.33	0	0.33	0	-0.33
	0.2	-0.22	-0.14	-0.11	-0.12	0.18	-0.23	-0.13	-0.14	0.17	-0.24	-0.12	0.19	-0.23	0.18	-0.24	-0.59
	0.47	0.22	0.14	0.11	0.12	0.48	0.23	0.13	0.14	0.5	0.24	0.12	0.48	0.23	0.49	0.24	-0.08
Maximum																	
	0.42	1	0.93	0.99	0.22	1	0.82	1	0.99	0.91	1	0.72	0.98	0.97	0.34	1	0.79
	-10.52	4.48	6.54	3.95	11.29	-1.71	13.29	-2.58	4.75	-8.25	6.75	7.33	-5.67	9.33	-13	2	15
	-25.65	-20.99	-9.19	-8.7	-2.68	-19.03	-13.54	-17.66	-11.45	-27.42	-21.31	-5.89	-22.4	-17.12	-30.75	-25.11	-13.98
	4.62	29.96	22.27	16.6	25.25	15.6	40.12	12.5	20.95	10.92	34.81	20.56	11.06	35.79	4.75	29.11	43.98
Standard Deviation																	
	0.33	1	0.98	1	0.38	1	0.9	1	0.99	0.84	1	0.69	0.96	0.97	0.26	1	0.74
	-1.69	0.68	0.81	0.37	1.5	-0.58	1.78	-0.44	0.69	-1.39	0.97	1.13	-0.95	1.41	-2.08	0.28	2.36
	-3.96	-3.15	-1.55	-1.53	-0.6	-3.18	-2.25	-2.71	-1.74	-4.27	-3.25	-0.86	-3.47	-2.56	-4.75	-3.79	-1.99
	0.59	4.5	3.17	2.27	3.6	2.02	5.81	1.82	3.12	1.48	5.18	3.12	1.56	5.38	0.58	4.35	6.71
Percentile High Risk																	
	0.06	1	0.98	1	0.19	0.8	0.91	1	0.96	0.33	1	0.41	0.48	0.97	0.01	1	0.41
	0.21	-0.08	-0.08	-0.03	-0.16	0.13	-0.16	0.04	-0.09	0.2	-0.09	-0.13	0.16	-0.13	0.29	0	-0.29
	0	-0.44	-0.3	-0.21	-0.36	-0.12	-0.55	-0.17	-0.32	-0.07	-0.49	-0.32	-0.08	-0.51	0.04	-0.39	-0.7
	0.43	0.29	0.15	0.15	0.03	0.37	0.22	0.26	0.14	0.48	0.31	0.06	0.4	0.25	0.54	0.39	0.12
Alves Low Risk																	
	0.29	0.94	0.35	0.96	0.54	0.03	0.57	0.9	1	0.97	1	0.99	0.24	0.89	0.78	0.99	1
	0.13	0.12	0.13	0.05	0.1	0.21	0.19	-0.08	-0.03	0.08	0.06	0.05	0.15	0.14	0.11	0.09	-0.02
	-0.04	-0.18	-0.05	-0.09	-0.06	0.01	-0.12	-0.25	-0.22	-0.14	-0.26	-0.1	-0.04	-0.17	-0.1	-0.22	-0.35
	0.31	0.41	0.31	0.2	0.26	0.41	0.5	0.1	0.16	0.3	0.38	0.2	0.34	0.44	0.31	0.4	0.32
Alves High Risk																	
	0.98	0.97	0.02	0.04	0	0.04	0.24	0.99	1	1	1	0.98	0.99	0.99	1	1	1
	-0.03	-0.05	-0.1	-0.08	-0.11	-0.11	-0.13	0.02	0	0	-0.02	-0.03	-0.03	-0.05	0	-0.02	-0.02
	-0.12	-0.21	-0.2	-0.16	-0.19	-0.21	-0.29	-0.07	-0.1	-0.12	-0.19	-0.11	-0.13	-0.21	-0.11	-0.19	-0.2
	0.06	0.1	-0.01	0	-0.02	0	0.03	0.12	0.1	0.11	0.14	0.05	0.07	0.11	0.11	0.14	0.15
Alves Very High Risk																	
	0.77	1	0.02	0.08	0.99	1	0.9	0.97	0.21	0.1	1	0.66	0.37	1	1	0.99	0.94
	-0.03	0.01	0.06	0.04	0.01	0	0.04	-0.02	-0.04	-0.06	-0.02	-0.03	-0.04	0	-0.01	0.02	0.04
	-0.08	-0.07	0.01	0	-0.03	-0.06	-0.05	-0.07	-0.1	-0.12	-0.11	-0.07	-0.09	-0.09	-0.07	-0.06	-0.06
	0.02	0.1	0.11	0.08	0.06	0.06	0.13	0.03	0.01	0.01	0.07	0.02	0.02	0.09	0.04	0.11	0.13
Fan-out (FANOUT)																	
Alves High Risk																	
	0.91	1	1	0.98	0.09	0.94	1	0.95	0.12	0.9	1	0.5	1	0.96	0.98	0.43	0.91
	0.04	-0.04	-0.01	0.03	0.09	0.05	-0.04	0.04	0.1	0.06	-0.03	0.06	0.02	-0.07	-0.04	-0.12	-0.08
	-0.06	-0.21	-0.11	-0.06	-0.01	-0.07	-0.22	-0.06	-0.01	-0.07	-0.21	-0.03	-0.09	-0.24	-0.16	-0.3	-0.28
	0.14	0.13	0.09	0.11	0.18	0.16	0.14	0.14	0.2	0.18	0.16	0.15	0.13	0.11	0.08	0.06	0.11