Archetypal Personalities of Software Engineers and their Work Preferences: A New Perspective for Empirical Studies

Makrina Viola Kosti
Department of Informatics
Aristotle University of Thessaloniki
Thessaloniki, Greece
+30-2310991930
mkosti@csd.auth.gr

Robert Feldt
Department of CS & Eng.
Chalmers University
and Dept. of Software Engineering
Blekinge Institute. of Technology
Karlskrona, Sweden
+46-733-580580
robert.feldt@bth.se

Lefteris Angelis
Department of Informatics
Aristotle University of Thessaloniki
Thessaloniki, Greece
+30-2310998230
lef@csd.auth.gr

ABSTRACT

As the area of Software Engineering (SE) matures the role of human factors in software development is commonly recognized as important. Increasingly we see empirical studies that investigate the connection between, for example, personalities and preferences, attitudes or performances of software engineers. Statistical analysis holds a key role by providing the means for uncovering associations between various facets of human factors and behavioral effects on projects and outcomes. Traditional statistical techniques tend to explore and interpret the multidimensional personality and behavioral data from an "average-point" perspective, targeting central trends. This paper introduces a methodology with statistical tools that can provide a new and different perspective for this type of SE data. It seeks the boundaries of a psychometric dataset and discovers reference or "benchmark" personalities, the archetypal personalities. Then, the method examines the placement of all individuals in the dataset in relation to the archetypes. Furthermore, the SE preference characteristics, or generally, any other types of behavioral SE data, are analyzed with respect to the archetypes. As a case to exemplify the methodology we analyze personality and project preference data from 276 master level SE students and compare to previous "average-point" statistical analysis of the same data. We also discuss how Archetypal Analysis, the heart of the proposed methodology, combined with multi-correspondence analysis might be of general use in empirical SE.

Keywords

Archetypal Analysis, Software Engineering, Personality, MCA

1. INTRODUCTION

A significant part of Software Engineering (SE) research has been recently focused on various human factors. The impact of the human factors on the software industry and its projects is increasingly discussed and investigated in empirical studies (Hannay et al. 2010; Salleh et al. 2010b; Kosti et al. 2014; Lenberg et al. 2014). One of the reasons for the increasing interest and need for deep studies is that human factors, their interplay and their effects, are often complex, indirect and therefore significantly unpredictable. Despite progress in other areas of SE, e.g. technology and methods, the unanticipated character of human behavior can be the source of serious problems in software project management. Process choices, improvement procedures and various outcomes can become hard to predict, plan and manage. Furthermore, even though the effects of the individual's idiosyncrasies have been recognized already in the early development of the SE field (Olson 1980), relatively little empirical research has been devoted to better understand and utilize it in practice. We are only at the beginning of a long journey to collect, analyze and better understand the effects that human factors play in determining the success or failure of software engineering projects.

As any research area matures and starts collecting real-world data from surveys, case studies and experiments, there is an increased need for statistical techniques to analyze the data. This maturation has happened during the last 10-20 years within SE where a number of guideline papers, have been published (Kitchenham et al. 2009; Arcuri and Briand 2012), signaling the need for more systematic and rigorous collection, as well as for more advanced analysis of data, in order to tailor the gained knowledge for practical use. This becomes even more important when studying a multi-faceted, subjective and ambiguous area such as the impact of human factors in SE. Fortunately, the scientific areas of statistics, data mining and machine learning have made much progress in the last 30 years, and now provide more powerful tools to be applied in science and engineering, even when the investigated data are complex and interconnected. To realize the benefits the research communities need to evaluate and adopt the new statistical tools. An example can be found within the political sciences where Hainmuller and Hazlett (2013) recently proposed the new statistical technique of "Kernel Regularized Least Squares" for wider use within social sciences since it allows flexibility and power and at the same time ease of interpretability. By applying more modern statistical methods they show that a richer and deeper understanding within the social and political sciences can be created. We have similar aims for the analysis of SE data but based on another set of modern statistical tools.

Existing empirical research on human aspects in software engineering often focuses on finding connections between "soft" factors, such as personality, job attitude and performance on the one side, and preferences or project outcomes or effects on the other side. However, even when the latter have been operationalized and measured, the former "soft" ones are often complex and obscure concepts in the behavioral and social sciences making both theory building and data analysis complicated and challenging (Costa and MacCrae 1992). Possibly due to these reasons, i.e. the complexity and the nature of these multi-faceted connections, some studies tend to find clear associations (Feldt et al. 2010), while others find no or only small/few effects (Hannay et al. 2010). An overview of conflicting results can be found in our earlier paper (Kosti et al. 2014). Furthermore, for a comprehensive and systematic mapping study of research on personality in software engineering, we refer to a recent article by Cruz et al. (2015), where the researchers have concluded that there is a growing interest on the subject during the last few years towards different directions. They also point out the contradictory results so far, a fact that can stimulate further research.

One reason for the conflict can be different choices of psychometric constructs and metrics used in different studies. As an example, many SE studies investigating personality of software engineers have used the Myers-Briggs Type Indicator (MBTI) (Acuña and Juristo 2004; Chao and Atli 2006; Karn and Cowling 2006). Personality-focused SE research is an active area of research with many applications and there have, for example, been studies to predict performance on specific tasks (Da Cunha and Greathead 2007), building effective teams (Gorla and Lam 2004) or peers in Pair Programming (PP) (Sfetsos et al. 2009) and, more generally, the search for the best person for a specific IT job (Capretz and Ahmed 2010). However, the MBTI has scarce support in empirical Psychology, is by many considered dated and has been heavily criticized (Furnham 1996). In our previous studies we have thus measured personality based on the Five Factor Model (FFM) (Wiggins 1996) and in Feldt et al. (2010) we found significant associations between personality factors and the attitudes and preferences of 47 industrial software developers.

Another reason for conflicting results or lack of clear connections can be the choice of analysis method. Typically, the traditional statistical techniques used for summarizing or representing collected data and for inference from that data are governed by the notions of centrality and dispersion. The basic idea is to discover central points within the data and then analyze the overall dispersion from these central points. Consequently, in a study on personalities and professional behaviors, such an analysis would try to find the "average (or central) personality(ies)" and then base the entire analysis on that notion. This philosophy is fundamental in statistical methodologies like ANOVA. Moreover, the notion of association, or correlation, is measured and interpreted in a "global" manner, i.e. we typically seek for an overall association between variables and not correlations that may occur locally, in specific parts of the multivariate space. Statistical methods based on these principles are the de facto ones used, for example in Acuña et al. (2009) they use correlation and regression trying to explore the relationship between personality and team job satisfaction. Similarly, Salleh et al. (2010a) in the quest of exploring PP effectiveness in academic performance use the MANOVA statistical method. The advantage of traditional statistical analysis is that it can capture the main trends and differences in a data set. However, when the sample sizes are small or there are no strong or clear effects to be found they are at a disadvantage. More modern statistical methods often can give a more detailed analysis and a higher precision and work with smaller sample sizes.

In the present paper we investigate one set of modern statistical tools that have not seen much use in SE. As a case study we use empirical data from our previous research that investigated the association between personalities and work preferences of software engineers using ANOVA models (Feldt et al. 2010; Kosti et al. 2014). Specifically, a research question originating from these studies, concerned the working preferences and generally the professional behavior of the "extreme personalities", i.e. of individuals in a dataset that on one hand have divergent personalities and on the other they can be considered as key reference or benchmark points for all the individuals. While our previous studies (Feldt et al. 2010; Kosti et al. 2014) revealed representative personalities which were locally centered according to five personality factors, we here aim to use statistical methods that more clearly separate groups of personalities and thus can give more detailed results.

A statistical methodology which can help formalize and provide these types of answers is the not so well known "Archetypal Analysis" (AA) (Cutler and Breiman 1994). AA has been used in Porzio et al. (2006) to segment markets using extreme individuals and recently was used for benchmarking effort estimation models (Mittas et al. 2014). The basic methodology used is simple and follows three main steps: a) the identifications of reference individuals (the archetypes, in our case based on their personalities) b) the analysis of their features and c) the comparison of all the observed individuals with the reference individuals. To help reveal associations from the individuals to the behavioral data, here on software project and work preferences, we then apply Multiple Correspondence Analysis (MCA) (Abdi and Valentin 2007). MCA can create a low-dimensional representation of the associations between answers on multiple questions or measurements. Together, the previous strategy can provide a statistically grounded but visual overview of a complex, multivariate dataset. We argue that this is especially important when there are multiple dimensions of the data or when the studied concepts are less clear-cut, such as with measures related to individuals and subjectivity.

Overall, AA is a multivariate statistical technique that essentially explores the boundaries of a dataset in order to find a few extreme points which can approximately define the convex hull of the dataset, i.e. points which set the limits of the entire dataset. These points are called "archetypes" and can be used as reference or benchmark points for the entire dataset. The word archetype comes from the Greek word "archetypon" and as explained in the dictionary (Merriam-Webster Online

Dictionary 2008), is the original pattern or model of which all things of the same type are representations or copies. Therefore the aim of AA is to find “pure types” within a dataset, such that all the other data can be represented as combinations (convex, i.e. linear combinations) of these “pure types”. The exploration of the boundary offers an alternative perspective of the data, different from the traditional techniques which focus on the central or “average” points. Once the archetypes are found, all the other points of the dataset are expressed by their relevant placement with respect to archetypes. If the aim of our research is to investigate other characteristics (e.g. behavioral), the interest is focused on how these characteristics are distributed near the archetypes.

In order to illustrate the approach based on AA, we consider as a case scenario the analysis of personalities of individual software engineering students in relation to their programming preferences. Since we have previously analyzed this dataset with more traditional, statistical tools this allows us to compare the outcomes of the different analysis. The data we analyze are the responses to psychometric instruments, measuring personality according to the FFM, by a total of 276 graduate students in a Master of Science program in SE at a Swedish University. The FFM factors were measured using the freely available IPIP FFM item sets (Donnellan et al. 2006). The students also answered additional questions about their preferences in software development and project work. The measurements were collected via web surveys in two SE courses over three years (2010-2012) and involved students enrolled during a 5-year period (2008-2012). It is our strong belief that this subarea of research within SE, is ideal for applying AA since the archetypal personalities in a dataset, containing responses from human individuals, is a concept that can be perceived and interpreted in a quite elegant and intuitive manner. However, we also argue that the proposed methodology can find other uses as an analysis methodology for empirical SE data.

Section 2 of this paper briefly presents previous studies (subsection 2.1) that have investigated personality in connection to SE and software development. The same section (subsection 2.2) also provides the theoretical background and an introduction to AA. In Section 3, we describe the case study and the data we used, followed by the methodology of our analysis and the results of our statistical analysis in Section 4. In Section 4 we also describe MCA and interpret the extracted results. Finally, we conclude with Sections 5 and 6, stating the threats to the validity of our study and discussing the findings of our work with directions of future work.

2. BACKGROUND

Psychometric instruments have been used in previous studies in several research directions. In this section we summarize the related research in SE and software development and also we provide a quick introduction to the statistical technique of AA.

2.1 Studies on Personality and Software development

Years of studies in the personality psychology area have led to the description of personality by a set of traits, which is a set of attributes that harbor information about the general tendencies in how an individual thinks, feels and behaves (Sabini 1995). Many studies have tried to use these traits in the field of software development in order to find links between personality and SE activities and performance. For the sake of brevity, we will summarize these studies, categorizing them according to the strength or lack of strength of the connection they uncovered between personality and one or more SE factors/covariates.

To begin with, there have been studies where no connections were found (Chao and Atli 2006; Salleh et al. 2009; Hannay et al. 2010; Salleh et al. 2010a; Salleh et al. 2010b). More specifically, Chao and Atli (2006) in their study found no statistically significant connections between the quality of code developed by pairs of programmers and their personalities. Furthermore, Salleh et al. (2009) performed a study based on students using the FFM. They resulted that differences in personality did not affect the academic performance of students who programmed in pairs. Hannay et al. (2010), also by using the FFM, tried to investigate the effect of personality on PP performance, without however finding strong evidence. Similarly, Salleh et al (2010a; 2010b) regarding the effect of conscientiousness and neuroticism on PP, on both studies, did not find evidence for distinguishing the performance of paired students between different levels of either conscientiousness or neuroticism. Secondly, there are some studies where only some connections have been found, not too strong though. Bell et al. (2010) for instance, tested the impact of personality characteristics on individual performance within a team environment and no strong correlations were found.

Moreover, some research work showed only one specific connection (Dick and Zarnett 2002; Acuña and Juristo 2004; Karn and Cowling 2006; Capretz and Ahmed 2010). Acuña and Juristo (2004) in their study reported that assigning people to roles according to their capabilities and personality, improves software development. Capretz & Ahmed (2010) suggested that taking personality into account in the role assignment process increases the probabilities of the project's success. Additionally, Dick and Zarnett (2002) stated that personality traits are necessary in order to distinguish candidates that are talented in PP. Also, Karn and Crowling (2006) investigated the effects of personality on the performance of SE teams, using the MBTI personality types. The study demonstrated that teams can work satisfactorily despite significant ethnic, religious and personality differences between individual members.

Finally, there are studies that show more than one connection (Acuña et al. 2009; Sfetsos et al. 2009; Martínez et al. 2010; Feldt et al. 2010; Rehman et al. 2012; Kosti et al. 2014).. For instance, Martinez et al. (2010), working in the direction of role assignment, introduced RAMSET, a role assignment methodology that relates personality, abilities and software roles for the integration of SE teams. It applies sociometric and psychometric techniques through a fuzzy approach. The methodology is applied in SE courses (Martínez et al. 2011) and according to the writers it can improve the efficiency of the classroom teams. Recently, Rehman et al. (2012), using the FFM, found specific links between SE roles and FFM traits. Specifically they showed that software analysts should have Extraversion and Agreeableness as main personality traits. Software designers should be highly agreeable and open to experience. Software developers should be extroversive, open to experience and agreeable. Software testers should have openness to experience and conscientiousness and software maintenance engineers should have openness to experience and conscientiousness as dominant personality features.

From the personality test point of view, most of the previous mentioned studies converge to the use of the MBTI test (Karn and Cowling 2006; Capretz and Ahmed 2010; Martínez et al. 2011), a test which does not take into account the strengths of personality along different dimensions. This fact reduces the statistical analysis power, with the latter having a direct impact on the results of the analysis itself. Moreover, we observe that the existing empirical studies, that link personality to working preferences towards software development, are contradictory.

Regardless of which base data have been collected, different methods of statistical analysis can investigate different aspects and answer different questions. Traditional statistical methods, like the generalized linear models we used in Feldt et al. (2010) and Kosti et al. (2014), seek to find significant dependencies based on a "centralized" philosophy, i.e. dependencies that hold "on the average". However, there are questions which have to be addressed under different perspective. More specifically, in the studies we consider, the data points are individuals and the main variables are personality scores expressed as numerical measurements. It is therefore natural, besides our interest for the "average personality", to investigate the "extreme personalities", i.e. the ones that define the boundaries of our dataset. In fact, this work was motivated by such a question, i.e. to find the individuals representing the extreme personalities of our dataset and to explore the behavior of these individuals. The aforementioned questions are very well addressed by AA, a statistical methodology developed by Cutler and Breiman (1994), which combines mathematical elegance and meaningful and interpretable results. In fact, AA can be used for more thorough analysis of the entire dataset, in combination with other statistical techniques, since it provides a new perspective of all the individuals, according to their closeness to the extreme points. The mathematical principles of AA methodology are reviewed in the next Section.

2.2 Archetypal Analysis

The principal idea of AA is that any data point in a multidimensional space, defined by a set of numerical variables can be represented as a mixture of specific data points. These points are located on the boundaries of the dataset and more specifically on its convex hull (convex polygon for 2-dimensional spaces and convex polytope for multi-dimensional spaces). Therefore AA finds extreme points, the archetypes, on the dataset boundary that spans the whole space of data points. Cutler and Breiman (1994) presented the theoretical basis of the technique by defining the problem of finding archetypes as a nonlinear least square problem and present an alternating minimizing algorithm to solve it.

In our context, the data points are individuals represented by numerical vectors of personality scores. Our aim is first to apply AA on this dataset and identify the archetypal personalities. Then, all other individuals can be described with weights showing the closeness of each individual to each archetype. Since for the same individuals there are data from answers to questions related to work preferences when working in software projects, these answers are subsequently studied with respect to the archetypal personalities.

Formally, the mathematical problem, using terminology of our area of interest, can be described as follows: Consider a $n \times m$ matrix $\mathbf{X}$ representing a multivariate dataset with n observations of individuals and m personality attributes. For a given k value, the problem is to find a matrix $\mathbf{Z}$ of k archetypes, which are essentially m-dimensional vectors. Specifically, the goal is to find the two coefficient matrices $\mathbf{a}$ and $\mathbf{b}$ that minimize the residual sum of squares (RSS):

$$RSS = \left\| \mathbf{X} - \mathbf{a}\mathbf{Z}^T \right\|_2 \text{ with } \mathbf{Z} = \mathbf{X}^T\mathbf{b} \qquad (1)$$

where $\left\| \bullet \right\|_2$ denotes the Euclidean matrix norm, subject to the following constraints:

$$\sum_{j=1}^{k} \mathrm{a}_{ij} = 1 \text{ with } \mathrm{a}_{ij} \geq 0 \text{ and } i = 1,\ldots,n \qquad (2)$$

$$\sum_{i=1}^{n} \mathrm{b}_{ji} = 1 \text{ with } \mathrm{b}_{ji} \geq 0 \text{ and } i = 1,\ldots,k \qquad (3)$$

These constraints imply that the approximated data are convex combinations of the archetypes, i.e. $\mathbf{X} = \mathbf{a}\mathbf{Z}^T$, and also that the archetypes are convex combinations of the data points, that is $\mathbf{Z} = \mathbf{X}^T\mathbf{b}$. The term "convex combination" refers to the linear combination of points, when all coefficients are non-negative and sum to 1. The algorithm developed in Eugster and Leisch (2009) reduces RSS in eq. (1) successively. The user is able to determine the number of archetypes which efficiently represent the convex hull of the dataset. This is achieved by inspecting a simple "scree plot", i.e. a plot showing the reduction of RSS as the number of archetypes is increasing. Typically, we stop when the increase of the number of archetypes has minimum contribution to the further reduction of RSS ("elbow" criterion).

The a-coefficients of equation (2) play a very important role in the interpretation of the analysis using AA. They are the weights used for representing all data individuals as mixtures (synthesis), and specifically as convex combinations of the archetypes and essentially show the closeness of any individual to any archetype. Hence the individuals located very close to an archetype or identified as archetypes have a-coefficients equal or very near to 1 for the specific archetype and coefficients equal to 0 for all the other archetypes. We remind that the a-coefficients take values between 0 and 1 and their sum for any individual is 1. In our case study, we utilize the a-coefficients in order to investigate if and how closeness to certain archetypes is associated to specific software engineers' attitude/preferences.

As a method for analyzing and understanding the heterogeneity of software engineers (early career stage, as we use post-graduate students for our analysis) we use AA to investigate how these differences link to the software engineers attitudes/preferences. In order to facilitate the understanding of the method we present the following example, Fig. 1, deriving from the data we use in this study. These data are presented in subsection 3.1. You can find the meaning of each abbreviation in this figure in Table 1.

As we explain in the following sections, we found four archetypes in our dataset. In Fig 1 you can see some of the characteristics of one of these archetypes (here labeled Z1). Three of the individuals located close to this archetype were subjects with IDs 16, 88 and 192, having a-coefficients equal to 0.62, 0.55 and 0.81 respectively. Like all individuals close to archetype Z1, these individuals have high extraversion scores and relatively low conscientiousness scores. Regarding their working preferences/attitude choices, we observe that they agree with the choice of the archetype regarding working after a given schedule / project plan **(SchPlan)**. Always with respect to the choices of Archetype Z1, the remaining 3 preferences among the three individuals vary. Thus the archetypal analysis can help us capture overall patterns among individuals in our data while still allowing individual variation.

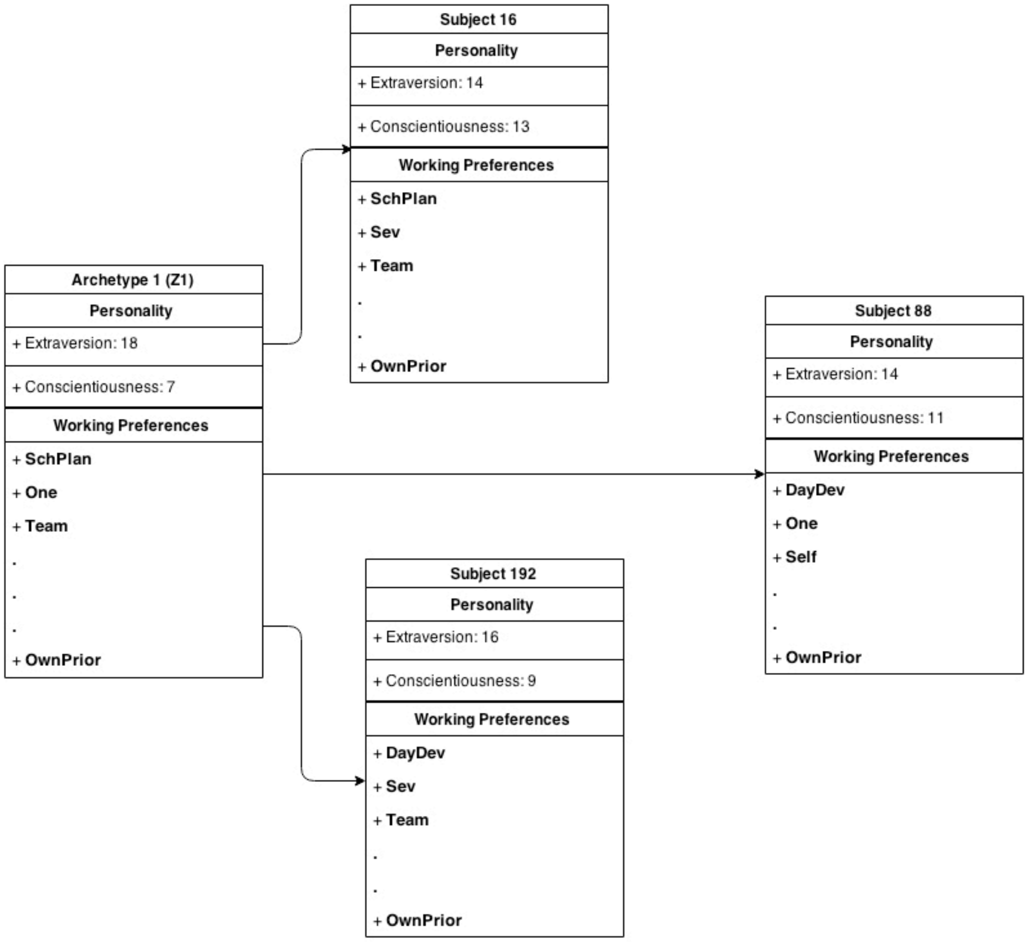

Fig. 1. Overview of proposed statistical analysis steps and results

3. CASE STUDY

In order to illustrate the new perspective that AA provides to the study of association between personalities and work preferences of software engineers, we used data collected from post-graduate students of a Master of Science program in SE at the Chalmers University of Technology and Gothenburg University in Gothenburg, Sweden. We were able thus to gather data in three overall measurements events, constituting a total of 276 unique subjects (students). All subjects answered the IPIP personality item (Donnellan et al. 2006) as well as a set of questions on their background, experience and SE preferences and attitudes.

Apart from the AA, which was the central statistical method in our research, we applied other statistical methods as well. Specifically, to be able to reveal differences between the different subject's attitude/preference answer levels and archetype coefficients, we used the Mann–Whitney non-parametric statistical test. We present the results per attitude question in Section 4. Additionally, aiming to achieve more insight in the relations of among our "extreme individuals" and their behavior or preferences, and in order to be able to represent how these characteristics are distributed near the archetypes, we used MCA. The explanation of this technique, followed by its results and interpretation can be found in Section 4.5.

3.1 Software Preference Questions and Overview of Answers

The data were collected in 3 different years using three different questionnaires each time, but each with a core set of questions that remained the same. With the purpose of statistically analyzing all respondents together, we merged the three datasets to one, using their common parts. Two out of the eleven common questions were of demographic nature while the remaining nine related to software development preferences. Additionally, the respondents were asked to complete the mini–IPIP psychometric questionnaire. We used the recently released, 20-item, mini version of IPIP (https://www.msu.edu/~lucasri/ipip.html), despite having used the larger 50-item version in previous research. This change was motivated by the fact that we posed several sets of questions to the students in this study and we needed to keep the total response time at a minimum. After data cleaning, 168 answers were used. This happened due to sparse missing answers to several questions by 108 students. Although the aforementioned students had answered to the IPIP psychometric questionnaire, they had missing answers to the part of the questionnaire regarding attitude/preference. To answer to the needs of our methodology, we removed these records from our dataset.

After defining the archetypes of our dataset, our goal was to find their associations to the working preferences. In order to have responses for all individuals to the questions regarding working preferences we removed all students that had not answered that part of the questionnaire.

Table 1. Descriptive statistics of the respondents who answered the IPIP questionnaire

iv. In your previous software development projects do you prefer to work:?	**(%)**	**v. Do you prefer working with:?**	**(%)**	**vi. Do you prefer working:?**	(%)
After a given schedule / project plan **(SchPlan)**	81.5	Several things at once **(Sev)**	36.9	In a team **(Team)**	82.7
As the day develops **(DayDev)**	18.5	One thing at a time **(One)**	63.1	By yourself **(Self)**	17.3
vii. Do you prefer to be responsible for:?	**(%)**	**viii. Do you prefer to work in:?**	**(%)**	**x. If you could choose would you prefer to work with:?**	**(%)**
Entire development process **(Entr)**	47	Longer projects lasting for several months up to a year **(Long)**	34.5	Technical parts of a software development project **(Tech)**	41.1
Particular part of development **(Part)**	53	Short projects lasting up to a couple of months **(Shrt)**	65.5	"Softer" / Management parts of a software development project **(Soft)**	58.9
ix. Do you prefer to work:?	**(%)**	**xi. You work best / most efficiently when:?**	**(%)**		
On project startup **(Strt)**	13.7	When a manager prioritizes your tasks **(MngrPrio)**	20.2		
From project start to project end **(StrtEnd)**	82.7	When you can prioritize your own tasks **(OwnPrior)**	79.8		
Short contributions as needed **(ShContr)**	3.6				

Of the 168 final respondents, 20.8% were female and 79.2% male. 53% were between 25-30 years old, 39.3 between 20-24 years old and 6.5% between 31-50 years old. 53.6% had 1-3 years of full time experience in software development, 23.2% had 3-5 years, 16.1% had less than 1 year and 7.2% had 5-20 years of full time experience. The remaining descriptive statistics for the rest of the questions and answers are presented in Table 1. These are exactly the same questions used in our previous study (Kosti et al. 2014).

4. METHODOLOGY AND RESULTS

To get a better understanding of our respondents and their software engineering preferences we will apply a number of statistical methods in consecutive steps. An overview of the combined methodology can be seen in Fig. 2, and also links to the sub-sections in which each type of analysis can be found. There are three basic steps: (a) find extreme individuals, (b) study their differing preferences/opinions in detail and (c) create overview of links from individuals to their preferences/opinions. While the 2nd step is a more "traditional" statistical analysis using hypothesis tests, the 1st step uses Archetypal Analysis (AA) to better separate the respondents into groups and the 3rd step uses MCA to visualize the full dataset and the results. The AA analysis differs from previous statistical analysis applied on software engineers since it focuses on the outer extremes rather than the average individuals, and the MCA complements the detailed, hypothesis tests; while the 2nd step gives details on which differences between individuals are significant, the overview map produced by MCA helps summarize results and can help explain and transfer them to students and practitioners. Below we present each analysis step in sub-sections 4.1 (AA), 4.2-4.4 (hypothesis tests) and 4.5 (MCA).

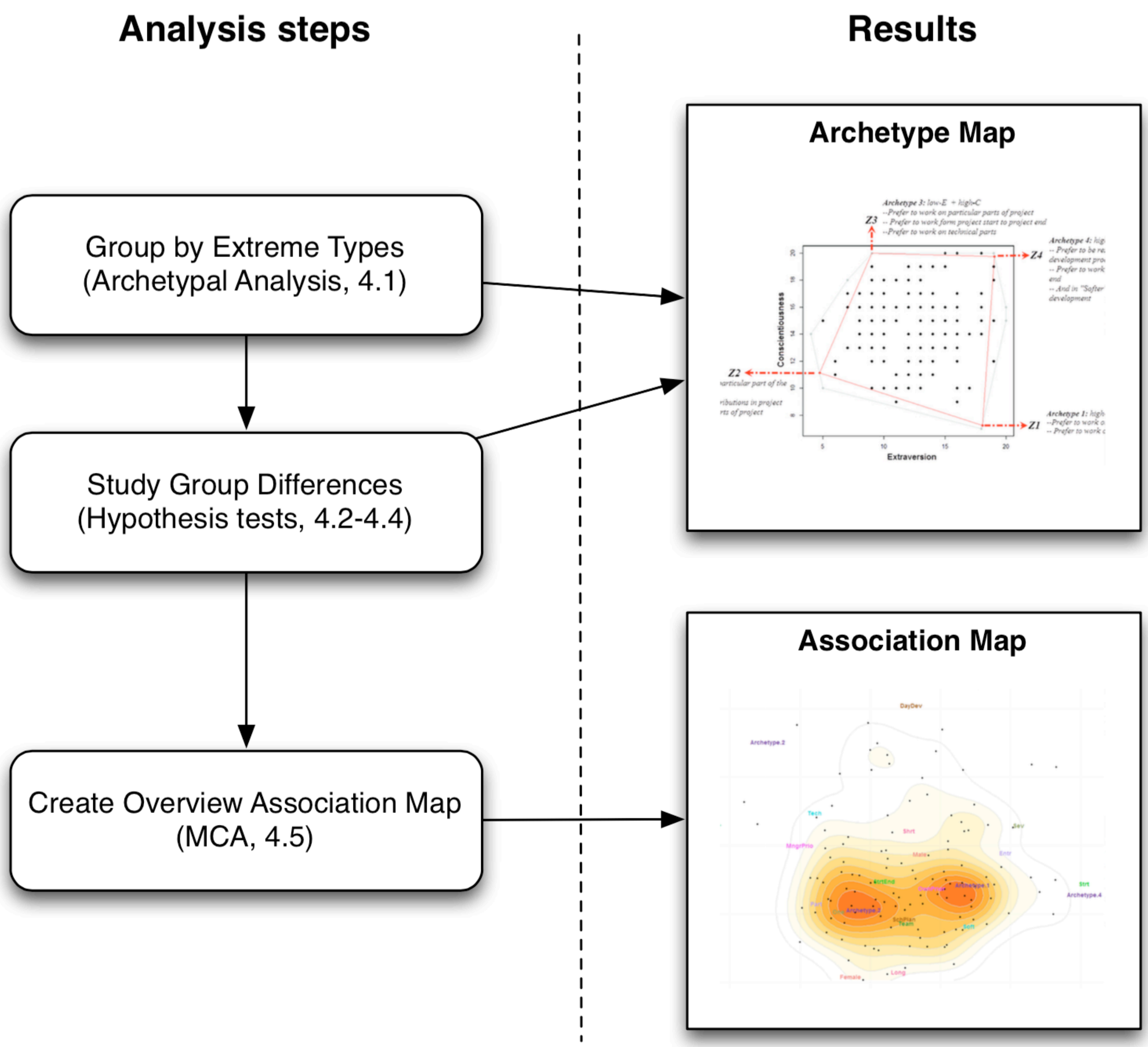

Fig. 2. Overview of proposed statistical analysis steps and results

4.1 Finding Extreme SE Individuals with Archetypal Analysis

In our case study involving master students, initially we have a 5-dimensional space for the individuals, described by the five FFM attributes used to describe the personality of our subjects. These attributes are: *Extraversion*, *Agreeableness*, *Intellect/Imagination*, *Emotional Stability* and *Conscientiousness.* In our previous work and during the preliminary phases of the analysis we noticed that these attributes exhibit a strong correlation pattern. In order to find more clearly separated archetypes it can often help to avoid including dimensions of the dataset that are highly correlated. Investigating this can be seen as an optional pre-processing step. It is not required by the archetypal analysis but can help improve its results. By improving results, we refer to the extracting of clearly differentiating archetypes. Also, if some dimensions can be excluded it also makes it easier to present the results; the data can be mapped in fewer dimensions.

For our dataset we performed this optional step using a traditional factor analysis (FA) with the principle component method of extracting the factors (Bartholomew 2002), which showed how the personality attributes are grouped. In Table 2, we can see the Rotated Component Matrix, i.e. the matrix of the FA factor loadings for each personality attribute onto each FA component. We observe that the five personality attributes are loaded in two groups (components). Component 1 is comprised of Extraversion (most strongly loaded onto the component), Agreeableness, Intellect/Imagination and Emotional Stability, while component 2 is comprised only of Conscientiousness.

Table 2. Rotated Component Matrix

	Component	
	1	2
Extraversion	0.746	
Agreeableness	0.698	
Intellect/Imagination	0.668	
Emotional Stability	0.466	-0.401
Conscientiousness		0.887

On the whole, the aim of AA is to find a few, not necessarily observed, boundary points (the archetypes) in a multivariate dataset such that all data can be well represented as convex combinations (linear combinations) of the archetypes. The archetypes themselves lie on the dataset boundary (the convex hull). Thus, archetypes can be seen as data–driven boundary values. Since the boundary of a multidimensional dataset is hard to be represented graphically for more than two dimensions, we have chosen the most strongly loaded attributes of the components from the factor analysis, i.e. Extraversion and Conscientiousness. Together, they efficiently capture the correlation structure of the entire dataset, and are related to our general purpose, that is to illustrate the position and the meaning of archetypes. In the two-dimensional space spanned by the selected attributes, the "shape" of data points, their convex hull, and the position of archetypes, can then more clearly be depicted. To perform these analyses we used the R statistical language (Team 2005) and the corresponding {archetypes} package that implements AA in R (Eugster and Leisch 2009). For the factor analysis we used SPSS statistical software (IBM SPSS 2013) but there are also R packages that can be used if one wants to stay with one analysis software, for instance the factanal() function of {stats} R package.

When performing an archetypal analysis one has to select the number of archetypes to use to describe the data. One good way of doing this is with a scree plot that plots the RSS of the AA versus the number of selected archetypes. One wants to select the least number of archetypes that captures most of the variation of the data set. In a scree plot this is often one of the "knee points", i.e. points which reduces the RSS the most. In our case, after applying AA, four archetypes were enough to minimize the RSS, as can be seen from the scree plot of Fig. 3. Specifically, the AA algorithm (Eugster and Leisch 2009) was executed for different numbers of archetypes and apparently the utilization of more than four archetypes does not contribute much to the reduction of RSS. As seen in Fig. 4, more than four data points are located in the convex hull, however the four points chosen in the end ("elbow" criterion) are considered to give a good approximation of the convex hull. A similar plot of the results when selecting three archetypes did not give as clearly separated and understandable archetypes. Depending on the number of dimensions selected this is data dependent but by iterating between selecting dimensions, performing archetypal analysis and producing scree plots the analyzer can often fairly easily select a good trade-off. The different quantitative RSS values obtained can also give a good indication of a sensible choice. In the end, any analysis has to trade off between the accuracy and understandability of the results.

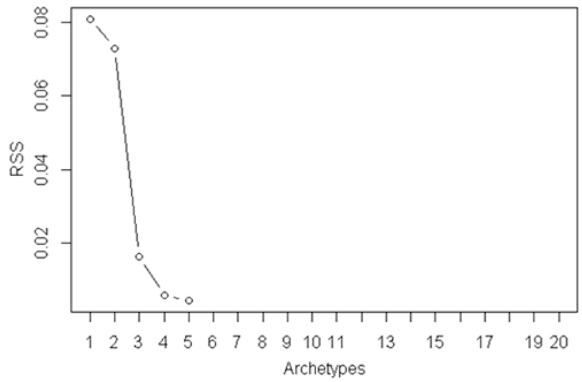

Fig. 3. RSS scree plot

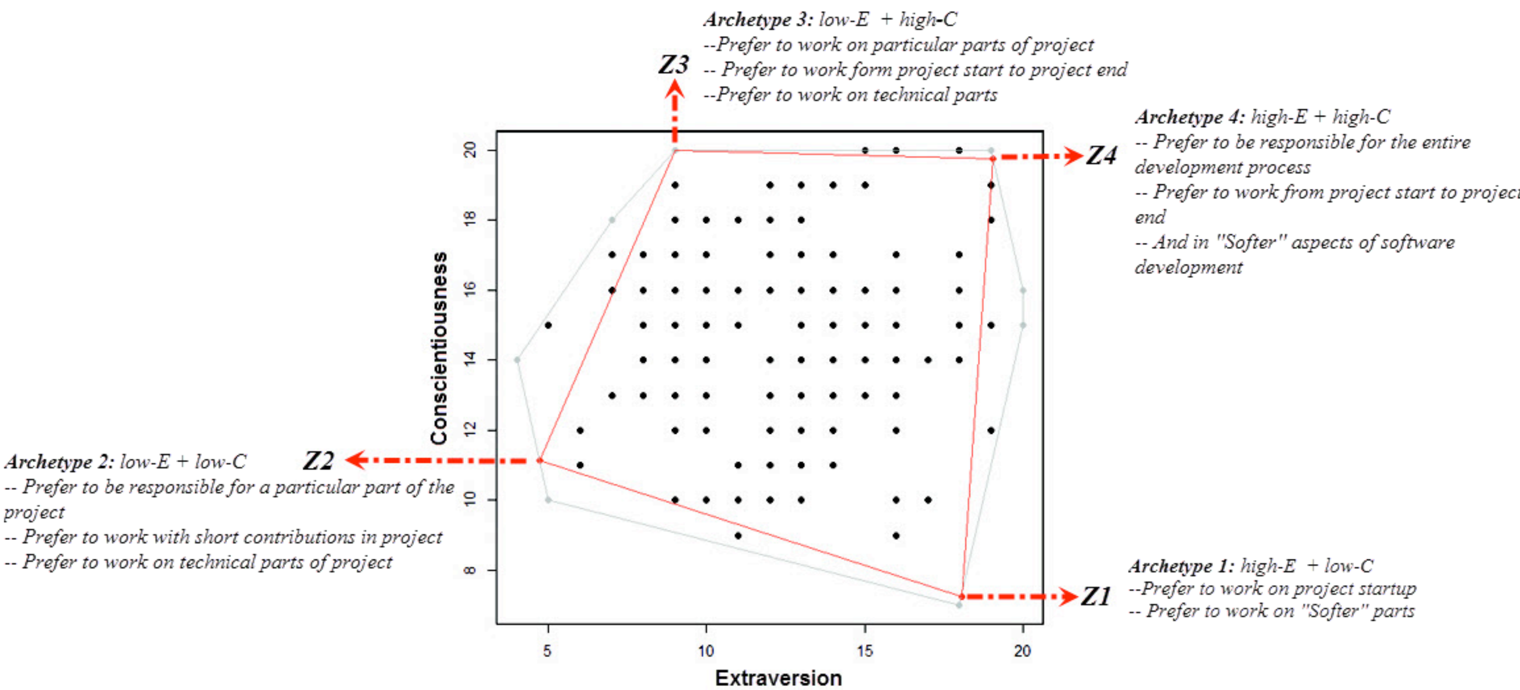

Fig. 4. Archetype Map with approximate convex hull (red lines) showing the 4 archetypes (Z1, Z2, Z3 and Z4) positions in relation to the other data points

In our case, the four resulting (synthetic) archetypes from our analysis matched four specific observations (students) of our dataset; see Table 3 (in bold are depicted the values of for the two personality factors for each synthetic archetype, while the non-bold faced values regard the ones of the matching observations of the actual dataset). In the following we will refer to these observations as, simply, (the) "archetypes". The scatterplot in Fig. 4 visualizes the differences between the 4 found archetypes and their relation to the levels of Extraversion (E) and Consciousness (C). The figure also summarizes the detailed results from the hypothesis tests performed in subsections 4.2-4.4 below; taken together we call it an "archetype map". We can see that in this case the four archetypes closely represent the four "binary combinations" of the 2-dimensional space:

high-E and high-C (archetype Z4), low-E and high-C (Z3), low-E and low-C (Z2) and high-E and low-C (Z1). This scatterplot also shows the approximation of the convex hull as defined from the positions of archetypes.

In AA the a-coefficients represent the amount of "similarity" of each observation to the corresponding archetype. For example, if $\alpha_2 = 0.8$ (for a specific non-archetypal observation), then the observation is 80% like Archetype 2 (Z2). In other words, the a-coefficients represent the amount of closeness of the observed individuals to a specific archetype. In the following subsections we present the results of the non-parametric statistical hypothesis tests (in our case we use the Mann–Whitney/Kruskal-Wallis tests, the non-parametric equivalents of t-test and ANOVA correspondingly) in order to reveal statistically significant differences between the different attitude answer levels and α-coefficients. These results are shown per attitude question, only for the questions where the test found statistically significant differences ($p \leq 0.05$) in subsections 4.2-4.4 below. The outcome of these detailed tests was then used to annotate each archetype in Fig 4 with their statistically significant preferences (in text bullets below each archetype). This archetype map is then further complemented with the association map based on the MCA in subsection 4.5 below.

Table 3. Synthetic archetypes calculated by the algorithm (with their closest specific observation in parenthesis)

	Z1 (obs. 205)	**Z2** (obs. 209)	**Z3** (obs.117)	**Z4** (obs. 194)
Extraversion	**18.06245** (18)	**4.717311** (5)	**8.999556** (9)	**19.06444** (19)
Conscientiousness	**7.250976** (7)	**11.13193** (10)	**20.0007** (20)	**19.7425** (20)

4.2 Project Responsibility Preferences

The test in this case showed significant differences for α_2 , α_3 and α_4 among the categories of question vii. (Table 2). The distribution of students for each one of the possible answers was 79-89, for "Entire development process" and "Particular part of development" correspondingly.

Regarding α_2 and α_3 , we observed that high levels of both coefficients were linked to preference in being responsible for a particular part of the development process ($p = 0.012$ and $p = 0.034$ respectively). Archetype 2 represents a student having low level of Conscientiousness and low level of Extraversion, while Archetype 3 represents a student with low Extraversion and high Conscientiousness. This means that students preferring to be responsible for a particular part of the development process could either be alike Archetype 2 or Archetype 3 (see Fig. 5 (a) for Archetype 3). About α_4 the test showed high levels of the coefficient when students showed preference of being responsible for the entire development process ($p = 0.002$). Archetype 4 (see Fig. 4) represents a highly extroversive and conscientious student. This means that individuals with these personality characteristics significantly prefer taking responsibilities in the entire development process.

4.3 Project Stage Preferences

The test accentuated four significant differences, linked to α_1, α_2, α_3 and α_4 among the three possible answers of question ix. (see Table 2). From a total of 168 observations, 23 preferred to work on project startup, 139 from project start to project end while the remaining 6 preferred to work with short contributions. Students near Archetype 1 and 4 show preference in working on project startup ($p = 0.033$ and $p = 0.031$ respectively). Archetype 1 represents a highly extroversive student with low Conscientiousness, while Archetype 4 highly conscientious and extroversive personalities. For α_1, see Fig. 5 (b).

Students near Archetype 2 seem to prefer working with short contributions. Archetype 2 represents a student with low levels of Extraversion and Conscientiousness. In contrast, students nearer Archetype 3 seem to prefer working from project startup to project end. Archetype 3, as seen in Fig. 4, represents highly conscientious students.

4.4 Work Type (Technical or Non-Technical) Preferences

The test revealed significant differences for all coefficients, α_1, α_2, α_3 and α_4. From a total of 168 observations, 69 students preferred to work in technical parts of software project while the remaining 99 preferred to work in "Softer" parts. Students near Archetypes 1 and 4 seem to have a preference in working in "Softer" parts of a software project. As we have already mentioned, Archetype 1 is characterized by high Extraversion and low Conscientiousness, while Archetype 4 symbolizes a highly extroversive and conscientious student. Students, who preferred working in technical parts of a software project, seem to be nearer to Archetype 2 and 3 than respondents preferring "Softer" parts (see Fig. 6).

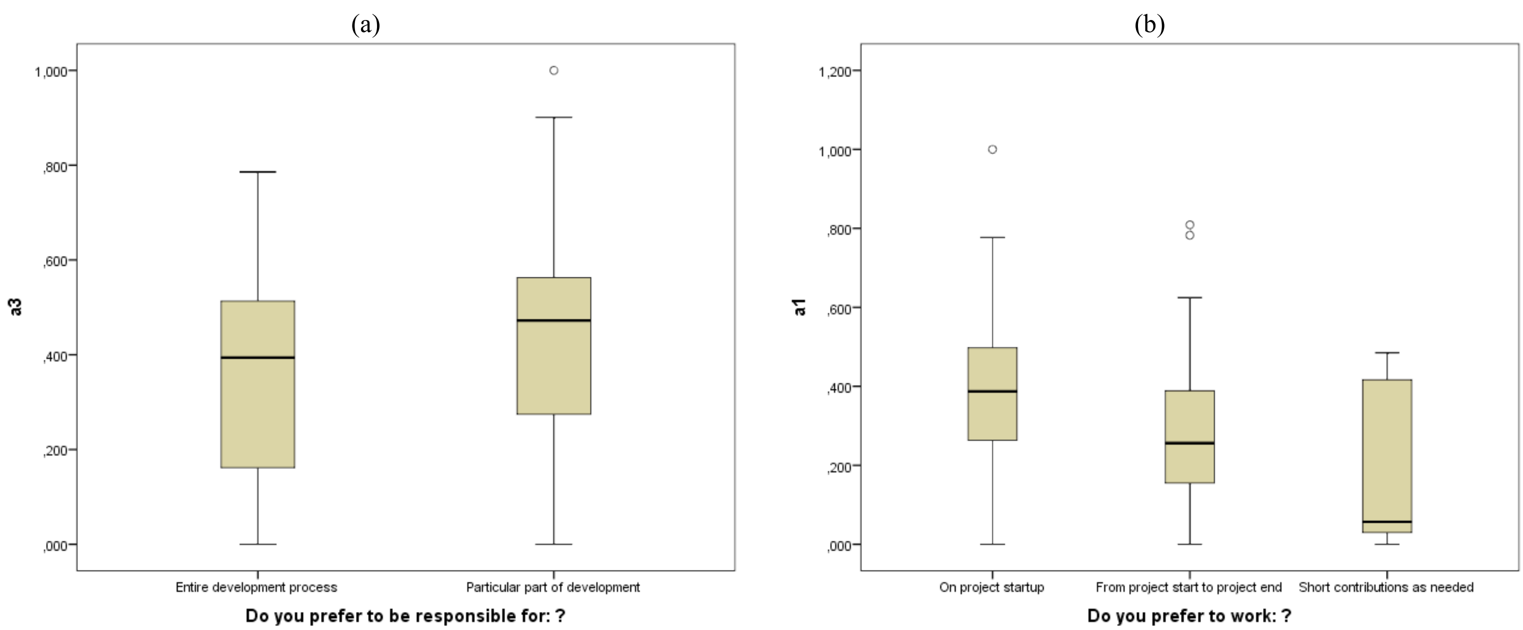

Fig. 5. (a) Boxplot with the distribution of question vii. in relation to Archetype's 3 coefficients. (b) Boxplot with the distribution of question ix. in relation to Archetype's 1 coefficients

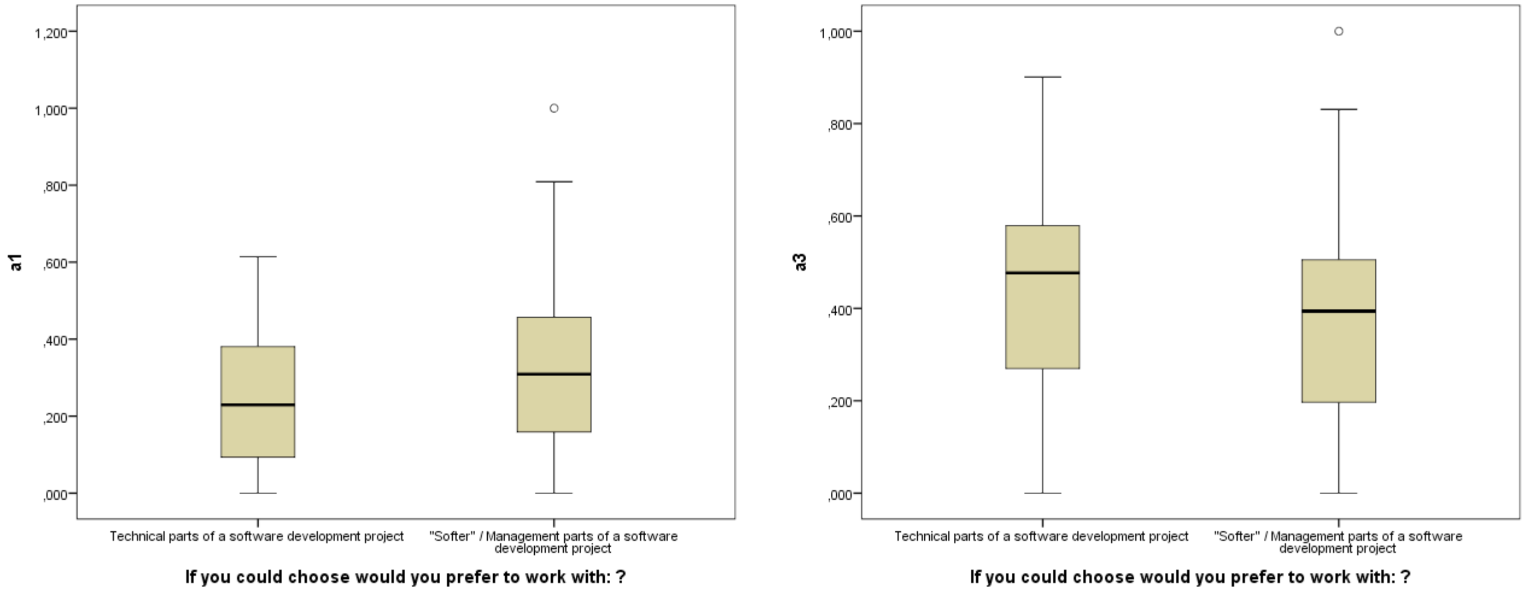

Fig. 6. Boxplots with the distribution of question x. in relation to archetypes' 1 and 3 coefficients (α_1 and α_3).

4.5 MCA and Interpretation of Results

The archetype map can summarize the archetypes found, the statistically significant associations between the archetypes and the preferences, opinions or other characteristics in additional data one has collected about them. Together they help alleviate the problems we saw in more traditional analyses of this type of data. However, with rich, multivariate datasets like these, we can further complement our results by using additional statistical methods and graphical representations.

MCA (Abdi and Valentin 2007) is an exploratory data analysis technique employed to detect and represent underlying structures in a dataset containing categorical data. This is achieved by representing multi-dimensional data points in a lower-dimensional space. Contrasted with traditional hypothesis testing, that aims to verify a priori hypotheses about relations between variables, exploratory data analysis is used to discover relations between variables when there are no a priori expectations of the nature of those relation and that would not be detected in a series of pair wise comparisons of variables. In a nutshell, MCA can show how the variables are related, especially on specific areas of the multidimensional space and not just that a relationship exists.

We chose MCA as it has several features that distinguish it from other techniques. For instance, as discussed above, MCA can uncover relationships that otherwise would not be detected in a series of pair wise variable comparisons. It can also be used on categorical data values which are common when questionnaires are used to probe a number of individuals. Thus, we do not need to map categorical answers of the individuals onto a real-valued, continuous scale. Moreover, it gives the opportunity to graphically display relationships in 2D plots, which is easier to interpret and can provide better understanding of the structure among the variable categories and objects (i.e., in our case the archetypes). The closer the categories are to the objects, the stronger the relation between that object to the specific category can be inferred. Finally, MCA is very flexible regarding data requirements. The only requirement is the existence of a rectangular data matrix with non-negative entries, or explained more simply, a $n \times m$ matrix, or, a data matrix whose size may not be the same in both dimensions. Many datasets can be represented or mapped into such a form.

Having explained the benefits of MCA we can now present the resulting two – dimensional MCA plot derived from our data (Fig. 7) using the {FactoMineR} package of R. The textual abbreviations inside the plot, pointing to the working preferences of our individuals, can be found in Table 2. They represent specific answers to the preference questions. For example, “One” refers to the answer “One thing at the time” to the question about which type of contribution the individual prefers to do in software engineering projects. The plot also shows the position of the archetypes. Textual items that are close to each other tend to be associated. For example, people that selected the answer “One” on the question above also tend to be close to Archetype 3. The heat map overlayed on the MCA-positioned textual categories shows the density of the individuals represented by individual black dots in the plot. Together, this “association map” thus shows both the associations and density of the individuals of the dataset; a rich representation of a high-dimensional dataset.

Specifically, the exploration of our data set with the archetypes and their working preferences shows a number of associations. The map in Fig. 7 indicates that Archetype 1 (highly extrovert individual with low conscientiousness) prefers working with “Softer” parts of the project development process (Soft) and prefers prioritizing their own tasks (OwnPrio). On the other hand, Archetype 3 (individual with low extraversion and high conscientiousness), strongly prefers working with one thing at a time (One), to be responsible for a particular part of development (Part), working on a team (Team) and after a given schedule/project plan (SchPlan) and also prefers working form project start to project end (StrtEnd). Moreover, we can see from the plot that near Archetype 4 there exist a number of individuals that prefer to work on project startup (Strt). However, we did not find any working preferences related to Archetypes 2. These results are somewhat different from the Archetype Map in Fig. 4 and show a limitation of MCA. Even though it can provide a better overview of the full dataset it is less accurate than the Archetype Map and the detailed hypotheses tests. Thus, we propose that neither should be used in isolation. By using both of them in a combined analysis we can get the positive elements of both.

The density curves we added to the MCA plots shows the highly concentrated zones of the observations, i.e. zones of the plot that contain dense accumulation of individuals. This way, we can explore if there are concentrations of individuals near the archetypes since the zones are colored according to the amount of density of individuals. Density variability changes from white and yellow (low density) to red (high density). For our dataset, we observe two formations with individuals centered near Archetypes 3 and 1.

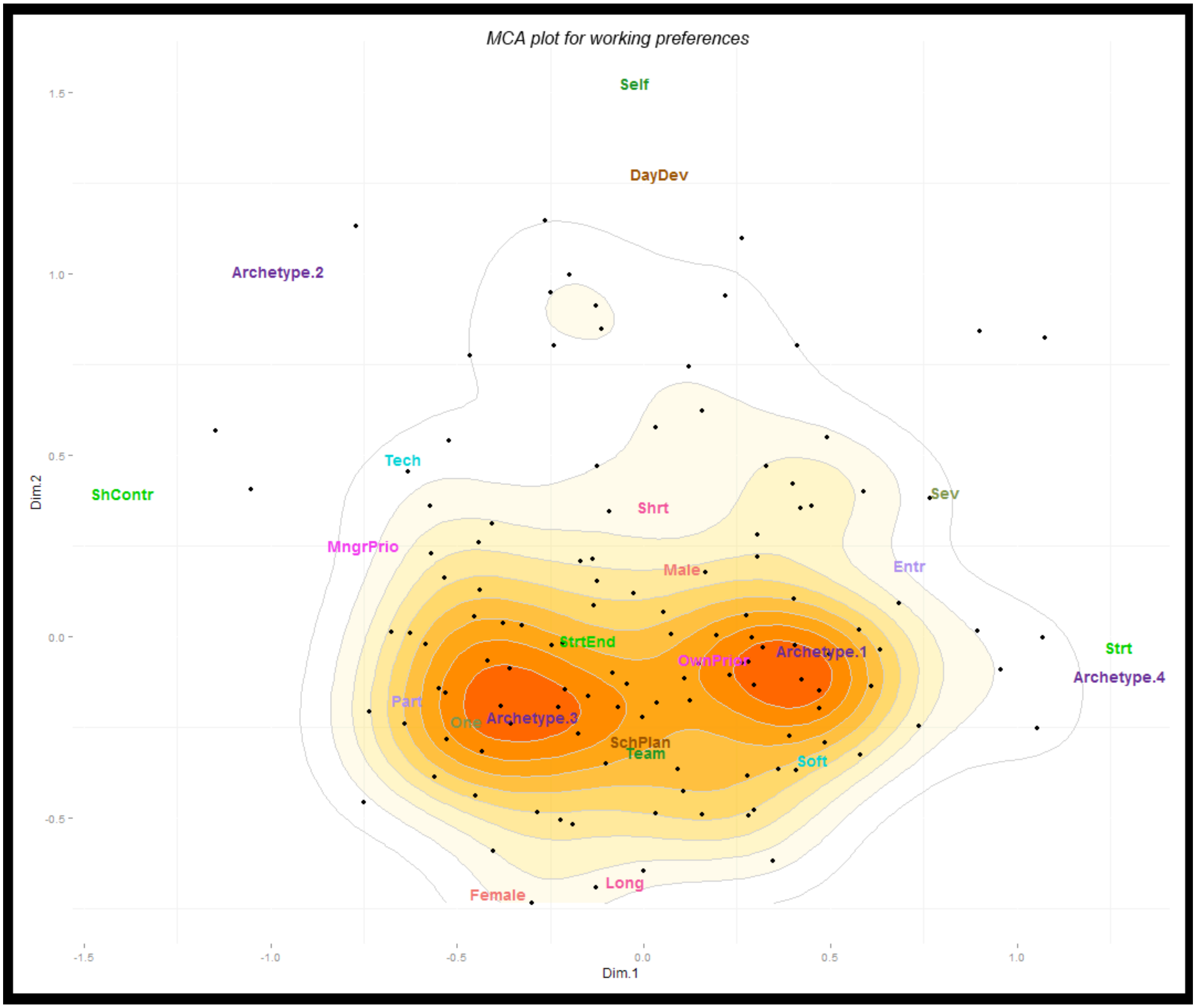

Fig. 7. MCA plot for the distribution of working preferences with regard to the archetypal students.

From the created zones we observe that the combinations of answers Part – One – StrtEnd – SchPlan – Team is more likely to come from Archetype 3 personalities. Moreover, the combination of answers OwnPrio – Soft defines Archetype 1 SE student types. From Table 1 we can also see the number of individuals of these two archetypes among the 168 individuals used in the MCA. We can only speculate as to why these two personality types are the most prevalent among the SE students. In a sense, archetype 3, can be interpreted as a classic "hacker stereotype" with an individual that is relatively introvert but focused on details. But we also see the opposite type among the responses; SE students that are relatively highly extrovert but less focused on details and more focused on "softer" aspects of SE. This main axis of our data thus supports our previous analyses (Kosti et al. 2014). The AA and MCA together provide evidence that there is a perpendicular axis among SE students that have other preferences. Together, this analysis thus presents a richer picture of the responses then our previous analyses did. We have also found the visual overview provided by the MCA plot to be useful when presenting the results and creating deep and detailed discussions about them both with SE students and colleagues.

5. THREATS TO VALIDITY

In this study we used graduate students. This helped us to have a good response rate, since the contact with the students was more direct than using professionals for the same scope. A small percent of students did not answer at all or their responses were not included, due to technical problems in the process of data collection. The total number of respondents used in the end is quite large and so in this regard there is no serious threat.

Regarding the psychometric instrument, we used the short version comprised of 20 questions. Although the short version has been constructed to include the most important items, based on extensive empirical studies within the field of psychology, this choice might have affected our results making it harder to uncover associations in the statistical analysis part. Fewer questions lead to less information so that individuals that are actually different may look similar in our data. This can make subsequent analysis harder. However, we would not have been able to conduct the study using the longer form of the IPIP questionnaire and we would risk that subjects got tired while answering the questionnaire and thus would not answer truthfully. Thus we find this a needed and tolerable trade off.

Furthermore, a threat to our study is that the studied SE preference questions do not give a detailed picture of the many preferences a software engineer actually has. The final preference questions came as a result of merging data collected during three years, in which different questionnaires were used at different points in time. Thus there is a risk that preferences might have changed slightly over the years. Research in psychology generally finds that peoples' opinions are more stable in time than not. However, it is clear that a more detailed set of preference questions should be used in future but similar work.

Selection of subjects - The fact that we use young (a majority of our subjects were born in the 1980's), graduate students with a homogenous age span impedes us from claiming the sample to be representative of the whole group of software engineers. Another constraint of our population has to do with the origin of the subjects. Even though the master programs have a large number of international students, the majority still remains Swedish which might introduce some bias.

Experimental mortality- The size of the full set of subjects was 276. But due to the multi-year data collection and the fact that some students leave university, switch programs or swap courses we did not have a full set of answers from all of them. Furthermore there were some with only some missing answers, maybe because of problems with the questionnaire itself or because they got bored. In the end a total of 108 subjects were excluded since we did not have complete data for them. This loss of subjects might mean that there is a potential bias in our data if the remaining subjects were more motivated to complete the entire questionnaire or if there is a pattern in which personality types or set of preferences that are more or less likely to conclude a university program in software engineering. We see no realistic way we can alleviate this threat.

Regarding factors that jeopardize external validity, as psychometric studies are based on self-assessment, evaluation apprehension might be an issue. The need of humans to "look good" and "smart" at some extent affects the sincerity of their answers. This is a common problem with personality tests, which are based on self-assessment. The risk though is even greater in our study, since we investigate students under evaluation in the end of the course. This adds an extra bias to our study. Although we took special care and ensured the students that we would keep anonymity to their answers and that the data would not be used for evaluation scopes, we cannot exclude the possibility that some of them were still afraid that their answers would affect their grades or progress. Thus, this might have led some of the students to give the answers that in their opinion their teacher expected them to give.

6. CONCLUSIONS AND DISCUSSION

Our results show that AA can provide a new perspective to empirical SE studies on personalities and work preferences or professional behavior. AA is able to identify typical individuals (the archetypes) and link their statistically significant preferences to groups of individuals that are close to these archetypes. This helps ground the results in existing responses/instances and makes it easier to convey and discuss the results. In particular this is helpful in teaching where students have less experience both with the subject area of SE, but also with statistical analysis and interpretation of its results in general, as in discussions with software practitioners and fellow researchers.

In our example case with personality data we found four archetypes that can in colloquial terms be described as the "meticulous extrovert" (Archetype 4), the "meticulous introvert" (3), the "relaxed extrovert" (1) and the "relaxed introvert" (2). Even though scientific analyses and presentations, of course, will need to stay away from simplistic categorizations we foresee that the pedagogical effects of providing more clear-cut metaphors and example cases on which to base a more detailed analysis can add value and create interest. This can have important effects both for education based on new empirical software engineering results but also in communication with industry practitioners.

In our previous studies linking personality of software practitioners to their software project preferences we used the statistical techniques of cluster analysis, factor analysis and generalized linear models (GLMs). Even though the latter, GLMs, gave results that could easily be interpreted in terms of the posed questions and measured personality factors, the former produced clusters that were hard for the practitioners to understand. When we used our results in the teaching of SE students even at the graduate (master level) they had a hard time interpreting the (non-GLM) results. AA alleviates these problems since the results can be related to specific responses and instances in the investigated data rather than to the more abstract idea of a cluster (from a cluster analysis), which is motivated from purely statistical reasons. Although, we have not yet systematically measured the pedagogical benefits from the AA results presented here our initial tests to present them both to master and PhD student level has indicated both that they are more easy to approach and "get into", and consequently, create more engaging discussions around the results. We plan to investigate this formally in future work.

Another benefit of the AA is that the results are more detailed. In previous studies we have found only two clusters of respondents, which have made the associations to the preference questions harder to analyze. The four archetypes identified in this paper lead to a clearer connections to the SE preference questions and support a conclusion that the personality factors of Extraversion and Conscientiousness are important when understanding and creating software development teams. A manager (or SE teacher) preparing for a new project could make use of this knowledge either to create a more aligned team (with similar preferences), or to create a more diverse and heterogeneous experience, which might be more robust or lead to "creative" discussions. Future work should investigate if similar patterns are also valid for SE practitioners; although several of the investigated students had quite extensive experience from professional software development it is still limited compared to engineers in industry. There can also be generational or age-related differences that should be explored (Kosti et al. 2014).

Since we have previously analyzed both this same data set (Kosti et al. 2014) as well as related data (Feldt et al 2010) we want to point out the differences with the approach we have used here. Both of our previous studies used more traditional statistical approaches with four main steps: (1) cluster analysis (to find any groups of subjects based on their personalities), (2) chi square analysis (to see if groups differ how they responded to SE preference questions), (3) ANOVA analysis (to find connections from each group to individual personality factors), and (4) detailed modeling of how personality factors connect to preference questions based on generalized linear models (GLMs).

The benefits of these analyses are that they are better established in the statistical literature. One can find extensive methodological descriptions of how to carry them out as well as example papers on their use. Conversely, although archetypal analysis has been around for some time it has seen relatively little use. It might have limits and biases that are not yet known although there is a growing body of evidence of its usefulness and fit-for-purpose. For example, Mörup and Hansen (2012) describe how "[AA] directly combines the virtues of clustering and the flexibility of matrix factorization".

Clustering is typically binary, i.e. an object is in one and only one cluster, which helps interpretability but is very inflexible, for example with objects that are in between clusters. Matrix factorization approaches such as SVD (singular value decomposition) and PCA (principal component analysis) are more flexible and can represent degrees of membership but typically lead to complex representations which are hard to interpret and thus do not lead to much insight. Since the chi square and ANOVA analyses we used in previous research depend on how the subjects have been assigned to clusters they suffer from the same inflexibility as the cluster analysis.

Using PCA or SVD might have solved the inflexibility but requires embedding that subjects in a space with scales that cannot be easily interpreted; we would see clusters but not be able to relate them to the actual personality measures. In contrast the archetype map in Figure 4 shows both the archetypes and the individuals on the original scale to which the researcher or practitioner can directly relate.

The GLM-based analysis we have previously used to numerically link individual answers on preference questions to individual personality factors give very detailed models that can be easily understood. However, they do not help to give an overview of the results, i.e. how individual answers to questions relate to each other or to the archetypes/clusters. This is the added benefit of the MCA and the association map. On the other hand, the association map does not give as precise an answer. When the increased level of detail is needed by a practitioner or researcher we thus propose that the analysis method we have presented here can be augmented by the GLM-based analysis from our earlier research. Together these two analysis types complement overview of the data with more detailed and interpretable models to quantify specific effects.

Compared to traditional cluster analysis and principal component analysis, we find that AA can better segment a dataset into separate groups; here we find four archetypes with clearly different characteristics, while our previous research focused on only two main groups. The archetypes can act in a way similar to personas as has been proposed in methods for requirements analysis (Aoyama, 2005). In fact, what we propose here can be seen as a data-driven way to derive personas which can then help structure and explain results. Most work on personas assumes that the personas are created subjectively, by experts or people with experience from the people being modeled. However, there have been proposals to infer personas from data collected about them and their behavior (McGinn and Kotamraju 2008). But they use factor analysis, which is based on finding the average behavior and clustering based on that. We argue that data-driven personas makes more sense when based on the extreme individuals being studied. This is what can be typically seen in market studies with the personas being based on age and coarse personality characteristics or extreme behavior. Thus, statistically a method like AA makes more sense; it takes the convex hull, i.e. the outer edge of the dataset as its starting point and bases the clusters based on that. Furthermore the archetypes are specific points in the space of the original data and thus can be directly used as the personas; with a traditional clustering analysis typically an average individual of the cluster must somehow be selected as the cluster summary. Furthermore, the combination of AA with MCA further adds the possibility of a single, visual overview of both the personas and their preferences and behaviors while providing more exact quantified data when details are needed. Future work should investigate if our methodology can help improve other areas of software engineering where personas have been proposed, such as in requirements engineering, usability and user experience work. Given our results it seems likely that AA can help give ad hoc methods for persona analysis a more sound statistical and theoretical grounding as well as lead to better results in practice.

AA is not a panacea though; our study uncovered several downsides. For one, AA in itself is not enough. As the methodology is data driven, we can supplement it with other methods which facilitate the simplicity of interpretation. For example, in our data, we discovered that there were strong correlations between personality factors, so we used that correlation to simplify the data and to be able to visualize the data and the archetypes in a two-dimensional space (Figure 4). The step however, is not compulsory and the original data can be used in order to decide if a factor analysis is actually needed. The main methodology (i.e. AA with MCA) can be applied without this step. Furthermore, the analysis presented here does not give as direct a connection from the personality factor values and the answers to the preference questions as the analysis based on generalized linear models of Feldt et al. (2010). However, the latter has a different focus (the questions rather than the clusters or typical cluster representatives) and thus could complement AA; there is no conflict between the two. A downside with actually executing an AA is that it is often not clear how to select the number of archetypes. However, most packages for AA have support for and implement validation procedures that help the selection of the number of archetypes.

To conclude, we propose that AA has many uses in the analysis of personality and preference data and in particular since its results can be more easily interpreted in terms of the investigated individuals. This has benefits in the education of SE students but the benefits are likely to be similar in conveying research results to industrial practitioners. Our specific results on 276 SE students can also be used as a baseline for future research on SE education. In particular, the personality factors of Extraversion and Conscientiousness seem to have an effect on how students prefer to work in software development projects. Future work should further investigate this as well as investigate associations also with project and examination outcomes. It should also study the application of AA and MCA to other areas of software engineering.

7. Acknowledgments

We acknowledge the support of Swedish Armed Forces, Swedish Defence Materiel Administration and Swedish Governmental Agency for Innovation Systems in the project 'Aligning Requirements and Verification Practices in Air Traffic Control Systems' (project number 2013-01199).

8. REFERENCES

[1] Abdi, H., & Valentin, D. (2007). Multiple correspondence analysis. *Encyclopedia of measurement and statistics*, 651-657.

[2] Acuña, S. T., & Juristo, N. (2004). Assigning people to roles in software projects. *Software: Practice and Experience*, *34*(7), 675-696.

[3] Acuña, S. T., Gómez, M., & Juristo, N. (2009). How do personality, team processes and task characteristics relate to job satisfaction and software quality?. *Information and Software Technology*, *51*(3), 627-639.

[4] Arcuri, A., & Briand, L. (2012). A hitchhiker's guide to statistical tests for assessing randomized algorithms in software engineering. *Software Testing, Verification and Reliability*.

[5] Aoyama, M. (2005, August). Persona-and-scenario based requirements engineering for software embedded in digital consumer products. In *Requirements Engineering, 2005. Proceedings. 13th IEEE International Conference on* (pp. 85-94). IEEE.

[6] Bartholomew, D. J. (Ed.). (2002). The analysis and interpretation of multivariate data for social scientists. CRC Press.

[7] Bell, D., Hall, T., Hannay, J. E., Pfahl, D., & Acuna, S. T. (2010, May). Software engineering group work: personality, patterns and performance. In *Proceedings of the 2010 Special Interest Group on Management Information System's 48th annual conference on Computer personnel research on Computer personnel research* (pp. 43-47). ACM.

[8] Capretz, L. F., & Ahmed, F. (2010). Making sense of software development and personality types. *IT professional*, *12*(1), 6-13.

[9] Chao, J., & Atli, G. (2006, July). Critical personality traits in successful pair programming. In Agile Conference, 2006 (pp. 5-pp). IEEE.

[10] Costa, P. T., & MacCrae, R. R. (1992). Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO FFI): Professional Manual. Psychological Assessment Resources.

[11] Cruz, S., da Silva, F. Q., & Capretz, L. F. (2015). Forty years of research on personality in software engineering: A mapping study. *Computers in Human Behavior*, *46*, 94-113.

[12] Cutler, A., & Breiman, L. (1994). Archetypal analysis. *Technometrics*, *36*(4), 338-347.

[13] Da Cunha, A. D., & Greathead, D. (2007). Does personality matter?: an analysis of code-review ability. *Communications of the ACM*, *50*(5), 109-112.

[14] Dick, A. J., & Zarnett, B. (2002, May). Paired programming and personality traits. In *Third International Conference on eXtreme Programming and Agile Processes in Software Engineering, Alghero, Sardinia, Italy*.

[15] Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The mini-IPIP scales: tiny-yet-effective measures of the Big Five factors of personality. *Psychological assessment*, *18*(2), 192.

[16] Eugster, M., & Leisch, F. (2009). From Spider-man to Hero-archetypal analysis in R.

[17] Feldt, R., Angelis, L., Torkar, R., & Samuelsson, M. (2010). Links between the personalities, views and attitudes of software engineers. *Information and Software Technology*, *52*(6), 611-624.

[18] Furnham, A. (1996). The big five versus the big four: the relationship between the Myers-Briggs Type Indicator (MBTI) and NEO-PI five factor model of personality. *Personality and Individual Differences*, *21*(2), 303-307.

[19] Gorla, N., & Lam, Y. W. (2004). Who should work with whom?: building effective software project teams. *Communications of the ACM*, *47*(6), 79-82.

[20] Hainmueller, J., & Hazlett, C. (2013). Kernel Regularized Least Squares: Reducing Misspecification Bias with a Flexible and Interpretable Machine Learning Approach. *Political Analysis*, mpt019.

[21] Hannay, J. E., Arisholm, E., Engvik, H., & Sjoberg, D. I. (2010). Effects of personality on pair programming. *Software Engineering, IEEE Transactions on*, *36*(1), 61-80.

[22] IBM SPSS software, http://www-01.ibm.com/software/analytics/spss, last access, July 20th, 2013

[23] Karn, J., & Cowling, T. (2006, September). A follow up study of the effect of personality on the performance of software engineering teams. In *Proceedings of the 2006 ACM/IEEE international symposium on Empirical software engineering* (pp. 232-241). ACM.

[24] Kitchenham, B., Pearl Brereton, O., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering–A systematic literature review. *Information and software technology*, *51*(1), 7-15.

[25] Kosti, M. V., Feldt, R., & Angelis, L. (2014). Personality, emotional intelligence and work preferences in software engineering: An empirical study. *Information and Software Technology*, 56(8), 973-990.

[26] Lenberg, P, Feldt, R. & Wallgren, G. L., (2014). Towards a Behavioral Software Engineering. In Proceedings of the *7th International Workshop on Cooperative and Human Aspects of Software Engineering*, June 2-3, Hyderabad, India, (pp. 16).

[27] Martínez, L. G., Rodríguez-Díaz, A., Licea, G., & Castro, J. R. (2010). Big five patterns for software engineering roles using an ANFIS learning approach with RAMSET. In *Advances in Soft Computing* (pp. 428-439). Springer Berlin Heidelberg.

[28] Martínez, L. G., Licea, G., Rodríguez, A., Castro, J. R., & Castillo, O. (2011). Using MatLab's fuzzy logic toolbox to create an application for RAMSET in software engineering courses. *Computer Applications in Engineering Education*.

[29] McGinn, J. J., & Kotamraju, N. (2008, April). Data-driven persona development. In Proceedings of the SIGCHI Conference on Human Factors in *Computing Systems* (pp. 1521-1524). ACM.

[30] Merriam-Webster Online Dictionary (2008). Archetype. In Merriam-Webster Online Dictionary, Retrieved December 13, 2013, from http://www.merriam-webster.com/dictionary/archetype

[31] Mittas, N., Karpenisi, V., & Angelis, L. (2014, September). Benchmarking effort estimation models using archetypal analysis. In Proceedings of the *10th International Conference on Predictive Models in Software Engineering* (pp. 62-71). ACM.

[32] Mørup, M., & Hansen, L. K. (2012). Archetypal analysis for machine learning and data mining. *Neurocomputing, 80,* 54-63.

[33] Olson, M. H. (1980). Review of software: Human factors in computer and information systems, by Ben Shneiderman, Winthrop Computer Systems Series, 1980. *ACM SIGMIS Database*, 11(4), 21-21.

[34] Porzio, G. C., Ragozini, G., & Vistocco, D. (2006). Archetypal Analysis for Data Driven Benchmarking. In *Data Analysis, Classification and the Forward Search* (pp. 309-318). Springer Berlin Heidelberg.

[35] Rehman, M., Mahmood, A. K., Salleh, R., & Amin, A. (2012, June). Mapping job requirements of software engineers to Big Five Personality Traits. In *Computer & Information Science* (ICCIS), 2012 International Conference on (Vol. 2, pp. 1115-1122). IEEE.

[36] Sabini, J. (1995). Social Psychology. W.W. Norton & Co Inc., 2nd edition.

[37] Salleh, N., Mendes, E., Grundy, J., & Burch, G. S. J. (2009, October). An empirical study of the effects of personality in pair programming using the five-factor model. In *Proceedings of the 2009 3rd International Symposium on Empirical Software Engineering and Measurement* (pp. 214-225). IEEE Computer Society.

[38] Salleh, N., Mendes, E., Grundy, J., & Burch, G. S. J. (2010a, May). An empirical study of the effects of conscientiousness in pair programming using the five-factor personality model. In *Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering*-Volume 1 (pp. 577-586). ACM.

[39] Salleh, N., Mendes, E., Grundy, J., & Burch, G. S. J. (2010b, September). The effects of neuroticism on pair programming: An empirical study in the higher education context. In *Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement* (p. 22). ACM.

[40] Sfetsos, P., Stamelos, I., Angelis, L., & Deligiannis, I. (2009). An experimental investigation of personality types impact on pair effectiveness in pair programming. *Empirical Software Engineering*, 14(2), 187-226.

[41] Team, R. D. C. (2005). R: A Language and Environment for Statistical Computing. ISBN 3-900051-07-0. R Foundation for Statistical Computing. Vienna, Austria, 2013. url: http://www. R-project. org.

[42] Wiggins, J. S. (Ed.). (1996). The five-factor model of personality: Theoretical perspectives. Guilford Press.