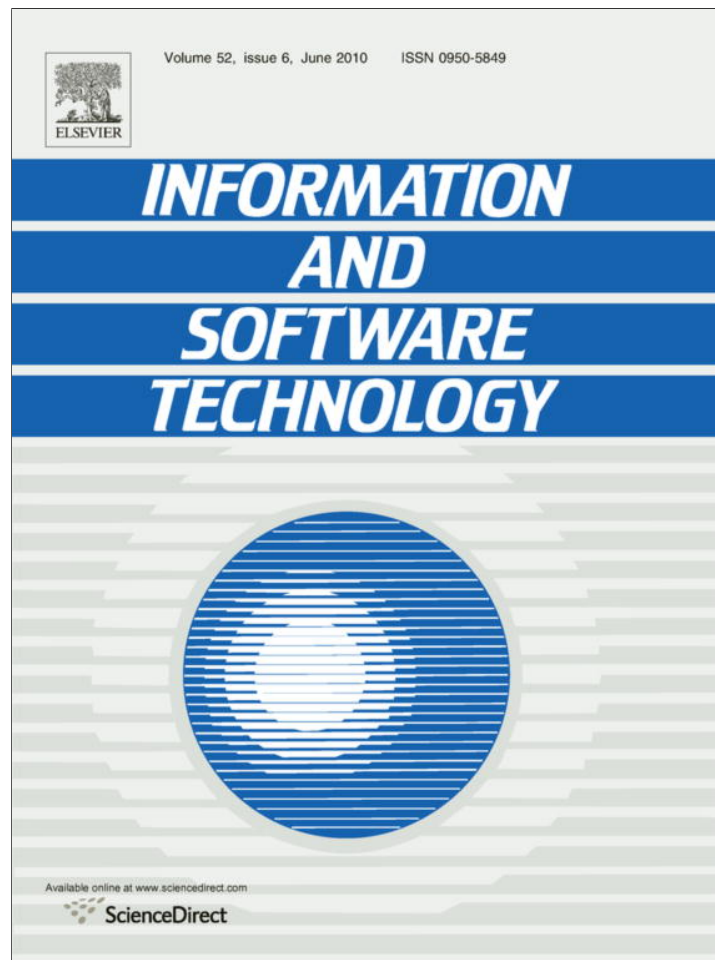


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

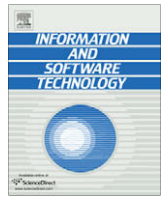
In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

Information and Software Technology

journal homepage: www.elsevier.com/locate/infsof

Links between the personalities, views and attitudes of software engineers

Robert Feldt^{a,*}, Lefteris Angelis^b, Richard Torkar^a, Maria Samuelsson^c^a Blekinge Institute of Technology, S-372 25 Ronneby, Sweden^b Dept. of Informatics, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece^c Dept. of Informatics, University West, S-461 86 Trollhättan, Sweden

ARTICLE INFO

Article history:

Received 8 November 2008

Received in revised form 11 January 2010

Accepted 11 January 2010

Available online 20 January 2010

Keywords:

Software engineering

Human factors

Personality

Attitudes

Empirical study

Statistical analysis

ABSTRACT

Context: Successful software development and management depends not only on the technologies, methods and processes employed but also on the judgments and decisions of the humans involved. These, in turn, are affected by the basic views and attitudes of the individual engineers.

Objective: The objective of this paper is to establish if these views and attitudes can be linked to the personalities of software engineers.

Methods: We summarize the literature on personality and software engineering and then describe an empirical study on 47 professional engineers in ten different Swedish software development companies. The study evaluated the personalities of these engineers via the IPIP 50-item five-factor personality test and prompted them on their attitudes towards and basic views on their professional activities.

Results: We present extensive statistical analyses of their responses to show that there are multiple, significant associations between personality factors and software engineering attitudes. The tested individuals are more homogeneous in personality than a larger sample of individuals from the general population.

Conclusion: Taken together, the methodology and personality test we propose and the associated statistical analyses can help find and quantify relations between complex factors in software engineering projects in both research and practice.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Software engineering is a broad field of study. In addition to the software itself and the many technical issues concerning its construction, development and maintenance, the field shares many aspects with most of the areas of modern business, e.g. project and product management, design, quality, customer expectations, legal issues, intellectual property and strategies. Thus, in order to develop effective knowledge on how to create successful software development projects we not only need to understand technical aspects but also in detail how the humans involved behave and why they make the judgments and take the decisions they do.

It has long been established that human performance and decision-making is affected by external factors such as stress [1] and in recent years internal factors such as cognitive ability and other individual differences have been in focus [2]. However, the connections between personality, job satisfaction and work performance are not simple with some results showing clear links while others do not [3–5]. To understand if, and how, individual differences affect judgment and performance in a particular field of study, such

as software engineering, we need specific results based on focused studies in the field in question. We also need methods to establish such results empirically.

Despite early interest in the importance of human factors in software development, and in particular personal characteristics of the humans involved in software engineering processes, such factors have been largely overlooked or not been based on empirical studies [6–10]. For example, even though a few studies have considered the personality of developers, they have used outdated models and metrics, that classify people into a few groups or types to do so [11–16]. This has prevented detailed knowledge about how different aspects of personality is linked to software engineering performance and hindered the use of this knowledge in improvement efforts.

This study aims to revive interest in using measures of human factors and personality to summarize, explain and predict the outcome of software engineering processes, methods and tools. We want to understand what is known about the connections between personality, decision-making and performance within software engineering. In particular, we want to see which personality models and metrics have been used and which problems they harbor. We then want to establish what are the alternative ways in which to measure personality differences, and then apply such, modern methods to measure and associate the personalities and attitudes

* Corresponding author. Tel.: +46 (0) 457 385 887.

E-mail addresses: robert.feldt@bth.se, rfd@bth.se (R. Feldt), lef@csd.auth.gr (L. Angelis), rto@bth.se (R. Torkar).

of professional software engineers. Taken together this leads to an empirically based method for establishing links between individuals and their software engineering preferences.

The aim of this paper is threefold:

1. To describe state-of-the-art personality models and tests that have a wide support within psychology that can be more generally applied in software engineering research.
2. To create and test a method for empirically studying links between personality and software engineering attitudes and preferences.
3. To provide extensive statistical analysis to find associations between personality factors and software engineering attitudes.

Section 2 of this paper provides more background on personality testing and the previous studies in software engineering that have used such tests. The design of our empirical study is then described in Section 3, followed by the results and statistical analysis in Section 4. Section 5 contains a discussion, while Section 6 concludes.

2. Personality, performance and software engineering

Personality is but one possible factor in explaining the judgments and decisions taken by humans involved in software development activities. Below we first put personality in context by describing the Blumberg and Pringle theory of work performance. We then discuss different personality models and metrics have been proposed to measure personality and its factors. Finally, we describe the previous studies linking personality and software engineering.

2.1. Theory of work performance

The judgments and decisions taken by humans while performing a work task to a large extent determines the quality of the performance. Even though personality is one factor in affecting these decisions, and thus in determining performance, there are a multitude of other factors that can affect performance. To put personality in context we briefly outline the model of work performance described by Blumberg and Pringle [17,2].

Blumberg and Pringle noted that there were a number of conflicting theories for explaining differences in work performance. The theories could be grouped as mainly focusing on the *Capacity* of an individual to perform the task or on the *Willingness* to actually perform it. However, they noted that there was a missing dimension; the performance on a work task also depends on the *Opportunity* the individual has to actually perform the task. The actual performance is determined by all three of these dimensions:

- Capacity (C)—The physiological and cognitive abilities of the individual that enables her to perform a work task in an efficient way,
- Willingness (W)—The psychological and emotional characteristics that influence the degree to which the individual is inclined to perform the task, and
- Opportunity (O)—The particular configuration of the environment surrounding, and beyond her direct control, an individual and her task that enables or constrains her performance.

Examples of capacities are the intelligence, skills, level of education, age, physical health and energy level of an individual. Without them the individual would not be able to perform the task, and the degree to which they are present determines to what level of quality the task can be performed. Within the Willingness dimension motivational and attitudinal factors such as motivation, job satisfaction, job status, self-image and norms are typical examples. Per-

sonality is also part of the Willingness dimension. Finally, in the Opportunity dimension, examples of factors are tools, materials, working conditions, leader behavior, rules and behavior.

Given this model it is clear that personality is but one variable in building our understanding on how human factors affects software engineering performance. Even though recent studies have found ample previous work in some other areas, e.g. the motivational models reported by Beecham et al. [5], for the long-term we see the need to investigate and relate multiple variables in different dimensions of the Blumberg and Pringle model.

2.2. Personality models and metrics

One of the main views within personality psychology is that personality can be described by a set of traits, i.e. fixed set of patterns in how a person behaves, feels and thinks [18,19]. These traits can be used to summarize, explain and predict how a person will act in different situations.

Even though different trait-based theories had been proposed since the 1930s it was not until the 1970s that they gained more widespread acceptance and interest. During this period many practical tests for personality traits were created and larger, empirical studies provided extensive data sets. There was much interest in applying these tests to match people to jobs and to put together successful and creative teams.

A major step was the Myers–Briggs Type Indicator (MBTI), based on Jungian theory, which to this day is the most used personality test [20,21]. It has four dimensions: Extraversion vs. Introversion, Sensing vs. Intuition, Thinking vs. Feeling and Judging vs. Perceiving. Based on 93 forced-choice items (only two options of which one has to be chosen), a licensed MBTI assessor can find the type of a person based on the largest score for each bipolar dimension. In theory, each of the 16 different personality types measured by MBTI can be viewed as collections of packaged traits.

However, the MBTI has been called into question by research in psychology since the late 1980s [22]. Empirical results have questioned whether MBTI measures qualitatively distinct types, and especially the Judging/Perceiving dimension seems weak. The alternative is a descriptive model called the Five-Factor Model (FFM) [22,23]. The five factors refer to broad personality dimensions that have been found in empirical research: Openness (O), Conscientiousness (C), Extraversion (E), Agreeableness (A), and Neuroticism (N). In recent descriptions the Neuroticism factor is more aptly called Emotional Stability [24]. In the following we will use the traditional term, but note that increased levels for the N factor actually refers to increasing emotional stability. The Openness factor is sometimes also called Intelligence but we will avoid that term since it can be confounded with IQ testing.

The FFM is a hierarchical organization of personality traits. Under each of the basic five factors there are six sub-facets for a total of 30 different personality facets [23].

There is ample support for that the five-factor model represents the five fundamental dimensions of personality [23]. They have been found in self-reports and ratings, by clustering analysis of adjectives in natural languages, in theoretically based questionnaires, in different ages such as for children, college students and older adults. Additionally, they have been found in samples of individuals from different cultures, such as English, Dutch, German and Japanese and they are also stable over different measurement instruments and observers and also across decades in adults [23]. Recently, some studies have found that there are even more general factors of personality that can explain some of the variation seen in the five factors [25]. However, the existence of even more general factors do not invalidate the importance of factors at lower levels and the five main factors and its sub-facets are the main levels of analysis in widespread use.

The NEO PI-R is the most used instrument to evaluate the personality of an individual based on the five-factor model [26]. It evaluates the full 30 facets of the FFM and takes around 40 min to take for one individual. There is also the NEO FFI version, which is scaled down to 60 items, and takes around 15 min to take for one individual. The problem with both of these instruments are that they are 'closed' and require licensed assessors with special certification before they can be used.

IPIP is a freely available set of items and scales for psychometrics based on the five-factor personality model [24,27]. The IPIP scales have several benefits compared to the MBTI: They have been found to better describe personality than the MBTI, they can be further refined to more detailed trait descriptions if needed (each factor is a 'super' factor collecting a number of sub-factors/traits together), they provide numerical scores for each factor, which makes more detailed statistical analysis possible, they do not require licensed assessors, and they are freely available. Like the NEO scales of the five-factor model they are available in different sizes depending on how much detail about each factor is needed.

2.3. Studies on personality and software development

There have been a few studies on the personality traits of software developers. Some have focused on programmers [12,28,15] and some, on the somewhat wider notion of software engineers [11,16]. All but one have used the MBTI to measure personality.

In [11], Capretz administered the MBTI to a sample of 100 software engineers. The main findings were that Intuition–Thinking and Sensing–Thinking personality types were over-represented in the sample, while Sensing–Feeling and Intuition–Feeling were underrepresented, compared to the general public. The distinction in the Thinking/Feeling dimension is concerned with how people make decisions; either following some logical sequence of facts (Thinking) or with more focus on feelings (Feeling) than logic. Capretz also summarizes previous, related research which seems to indicate that programmers are often introverts and that the types Introvert–Sensing–Thinking–Judging, Introvert–Intuition–Thinking–Judging and Introvert–Intuition–Thinking–Perceiving are the most frequent, while Extrovert–Sensing–Feeling–Judging, Introvert–Sensing–Feeling–Perceiving and Extrovert–Intuition–Thinking–Perceiving are underrepresented. In contrast to Capretz's findings, Smith found [14] that among 37 system analysts 35% were Introvert–Sensing–Thinking–Judging while 30% were Extrovert–Sensing–Thinking–Judging.

Recently, Chao and Atli [12] executed a personality survey with 60 programmer respondents. They found no evidence that there is a difference in code quality between different ways of pairing personality types in pair programming tasks. On the other hand, DaCunha and Greathead [15] found that students that were more intuitively inclined performed significantly better on a code review task of 282 lines of Java code. When considering pairs of MBTI dimensions together, students that were Intuition–Thinking types found two times more bugs than students who were Sensing–Feeling. Some studies have also used MBTI personality types to understand and improve software engineering education [13].

A recent systematic review by Beecham et al. [5], summarizes 92 papers with results on the motivation of software engineers. Forty-three of these papers were identified as relevant to the question of what are the characteristics of software engineers. 52% of the papers indicates that software engineers form a distinct group while 24% indicates that they do not, while 22% said that it is context-dependent. They identify 16 different 'raw' characteristics of software engineers of which growth orientation (wants to be challenged and learn new skills), introversion (low need for social interaction) and need for independence, are the most cited ones.

Acuna and Juristo is an exception to the other studies since they did not use the MBTI in their work on how to assign people to different roles in software projects [16]. Based on the 16 PF-5 personality test they proposed a model for software processes based on the capabilities of the involved people. With capability they mean a skill or behavioral attribute that can be related to 'activity-oriented behavior' in a software process. Their work surveys existing frameworks addressing human factors in software projects, such as People-CMM, Soft Systems Methodology and ALF, and is then based on the observation that the existing frameworks do not define the capabilities of people and roles and their interaction. Furthermore, Acuna and Juristo propose a 'table of correspondence' between the 16 personality factors and important software project capabilities. This is an original and important contribution; however, no empirical basis is given for the correspondence table. In this work we propose a methodology for establishing such correspondences empirically. This is also in contrast to previous work that have no empirical basis, such as [10].

Preliminary results from the study in this paper was presented in a short paper at a workshop [29]. That paper calls for a more widespread use of psychometric measurements based on questionnaires in any empirical software engineering study, but only presents a few associations for a single personality factor. This paper is a heavily extended version of the workshop paper and presents extensive analysis and discussion of the full data.

Common to many of the studies on personality of programmers or software engineers is that they have primarily focused on students and not on professional software engineers. Furthermore they have all used the now dated and criticized Myers–Briggs Type Indicator which classifies persons into different types and do not consider the strengths of their personality along different dimensions. This makes statistical analysis harder and can potentially hide associations between personality factors and the decisions taken and performance shown by software engineers. Also, the existing empirical material linking personality to views and attitudes towards software development and engineering are contradictory.

In summary, there is little empirical basis for claims of the importance of considering personality factors in software engineering research and practice. Also there is a lack of a methodology to establish such an empirical basis.

3. Design of empirical study

To study human and personality factors and their effect on software engineering we have chosen to do a web-based, standardized questionnaire. This should give a broader set of data suitable for statistical analysis that can indicate, in general, if personality factors are important to consider in software engineering research. In later stages, the trends and connections indicated by the statistical analysis can be used to steer the design of more qualitative studies such as deep interviews and on-site observational techniques. However, the existence of and the strength of personality factors must first be established with more certainty.

3.1. Questions and answer alternatives

The questionnaire had two main parts: A personality test (Part I) and a set of questions to probe the attitudes and working style and habits of the respondent in areas related to software engineering (Part II).

For the personality test in Part I we chose the IPIP 50-item scale. We chose the smallest scale with 50 items over the more detailed 100- and 300-item IPIP. That previous studies on software engineering and personality have mainly used students is probably due to the fact that it is hard to get hold of professional software

engineers; they have very little time available. In reason to maximize the possibility to conduct such a study in industry, we wanted to keep the total number of questions down. Thus we had to chose the smallest of the IPIP scales.

For each of the 50 items in the scale, the respondent should judge how accurately the item describes them in relation to their peers. The answer should be on a five-point scale: 'Very Inaccurate', 'Moderately Inaccurate', 'Neither Inaccurate or Accurate', 'Moderately Accurate', 'Very Accurately'. Respondents were encouraged to answer based on how they typically see themselves and not be too influenced by their current emotions and mental state.

Part II of the survey contained 56 additional questions on views, attitudes and working habits related to software engineering. It was divided into several subsections grouping related questions together. The first two subsections were of a general nature: Base data about the respondent (sex, age, years of experience), and about their software engineering roles and experience. Three additional subsections added more specific questions: software engineering tools and processes used, the developing organization, and the respondents view on software engineering research. Finally, a section on the respondents' self-image was added. Appendix A.1 summarizes Part II of the questionnaire in more detail.

Since one goal of the study was to make it repeatable, we tried to ensure that the wording of the questions made it hard for respondents to interpret them in more than one way, and that we only asked for one thing in each question. We added a free-text question at the end of the questionnaire where respondents could comment on the clarity of questions, answer categories and wording. The answer alternatives were mostly categorical. We chose to use, for most parts, Likert scales [30] with four alternatives, i.e. 'very low degree', 'low degree', 'high degree' and 'very high degree', to force respondents to take a stand. The exact wording of the answer alternatives changed slightly depending on the actual wording of the question but always had four alternatives corresponding to the ones above.

3.2. Pre-test procedure

Most of the questions had been previously used in a degree study by one of the authors; only eight new were added [31]. Thus we were already confident that the wordings of the questions and the sets of answer alternatives were suitable. Even so, a pre-test was performed with the help of a software developer from a company that did not participate in the study. The test was done to validate that the survey did not require too much time to complete and that the answers of alternatives were relevant for the questions and for the target respondents.

We explicitly designed the survey to require less than 40 min to take for all respondents since this was estimated to be the maximum time a commercially active software engineer would be willing to participate. The pre-test respondent needed only 19 min to complete the full survey. Even though he might not be representative this showed that a maximum of 40 min had likely been achieved. The pre-test respondent had a few comments on the wording of questions and they were refined together with him. He had no comments on the answer alternatives.

3.3. Ethical and integrity considerations

Since there are results linking personality measures based on the five-factor model with vocational interests and aptitude, the information gained from personality testing is potentially sensitive. For example, an employee might not want the employer to read the personality profile if that might negatively affect the employees career opportunities.

To overcome such fears and ensure that we had full control over the information from the respondents we opted to host the personality test on a computer at the university over which the researchers had full control. No other university employees could access the machine and once the results were gathered they were packed and encrypted and handled with utmost care. Furthermore we informed study participants that their answers would be available only to the researchers and not to anyone else in their company.

3.4. Target group and companies

The target group for the study was software engineers working in companies in Sweden that develop software for commercial use. The language used in the questionnaire, both in the introduction and the questions themselves, were Swedish.

The initial idea for the experiment was to get 20 respondents in three different software development companies to answer the questionnaire. However, after talking to a few initial companies, it became apparent that few of them would be willing to allow 20 employees to take the test. Most cited upcoming deadlines and time constraints as the reasons. However, some also showed some resistance towards "unproven methods" that are not common within software engineering or computer science. Instead we decided to contact a larger group of companies by phone.

When contacting the companies we asked to speak to one of the individuals responsible for the software developers at the company. In our first contact with the key persons at the companies, we explained the purpose and goal of our study. We also described how the raw data would be processed and that all answers of the respondents would be presented anonymously. In the second contact, when we dispatched the web survey, we reminded them that when the result was presented no respondent could be identified. We also informed them how to reach the web survey and the intended time frame for the survey. The information was also sent by e-mail to the companies.

A total of 10 software managers at different companies that matched our criteria agreed to send out information about the study to their software engineers. The information contained an introduction to the study with a link to the web survey and an invitation to participate. Seven of the companies are small- to medium-sized software development or software consultancy companies, one was a small subsidiary company within a large Swedish telecom company, one was a real-time and embedded software development department within a large industry, and one was the local branch of a large Nordic IT consultancy company.

4. Results and analysis

A total of 47 different respondents answered the web survey in two different time periods spanning 6 weeks during the summer of 2005. A number of different statistical methods were applied to analyze and find links between different aspects and factors of the responses, i.e. descriptive statistics and graphs, statistical tests, one way analysis of variance (ANOVA) and cluster analysis. We selected 28 of the 56 questions in Part II of the questionnaire for inclusion in our detailed analysis; primarily questions where the respondents could add and actually added additional answer alternatives were excluded. Finally, we applied a bootstrap technique for comparing the personality attributes of the software engineers in our study with relevant results from a previous published study. For all statistical test, a statistically significant difference is assessed when $p < 0.05$. However, since the tests are two-tailed, we report also the cases where $p < 0.10$ to indicate possibly significant differences or dependencies. Below we present our results.

4.1. Descriptive statistics of base data

Of the 47 respondents, 11 were female and 36 male. Fifty-five percent of the respondents were 24–35 years old, 17% were 36–40, and 23% between 41 and 55. None of the respondents were less than 25 years old and 4% were over 56 years old. These statistics are well in line with Swedish averages for the category ‘computer specialist’.

Eleven percent of the respondents had less than 2 years of software engineering experience, 19% 2–5, and 40% had 6–10 years of experience. Thirty percent had more than 10 years experience.

Forty-nine percent of respondents described themselves as foremost being ‘Programmers’, 9% as ‘System Architects’, 6% ‘Project Manager’, and 6% ‘Product Development Manager’. The rest have answered ‘Other’ and some further specified this as ‘Consultant’, ‘Team Lead’, or ‘Client Delivery Executive’. The results of the multiple-choice question on what kind of software engineering work the respondents did is shown in Table 1.

Very few free-text comments were given about the questionnaire itself by the respondents.

4.2. Cluster analysis of personality factors

Since each person’s personality is characterized by five attributes in the sample, it is interesting to investigate the existence of groupings in the data. Then we work with these groups in order to make our comparisons and find associations with the attitudes. The key point here is the use of all the personality attributes for the formation of the groups. In this way we consider all attributes together in our comparisons instead of testing each one separately. For this reason we used the multivariate statistical method, known as cluster analysis [32]. This method uses all the attributes together in order to identify homogeneous, mutually exclusive subsets (with individuals resembling each other) which, on the other hand, have large differences between them.

For our cluster analysis, we applied the SPSS algorithm ‘Two-Step Cluster Analysis’ (TSCA) on the five personality factors [33]. This specific algorithm was employed because we preferred not to fix a priori the number of clusters since there was no such prior information. The algorithm has a great advantage compared to traditional clustering techniques as it can automatically select the optimal number of clusters by comparing the values of a model choice criterion across different clustering solutions. After extensive experimentation with the various parameters of the algorithm, we used the AIC (Akaike’s Information Criterion [34] as our information criterion with the log-likelihood distance measure other options were tried but did not produce distinct clusters). The likelihood measure places a probability distribution on the variables, which are assumed to be normally distributed and independent.

To test our assumption of normal distribution we used the Kolmogorov–Smirnov test [35]. The five variables were tested for nor-

mality and we found that they did not differ significantly from the normal distribution ($p > 0.1$ for all variables). Regarding the assumption of independence, this is partly violated since according to Pearson’s correlation coefficient there is significant correlation ($p < 0.001$) between: (a) Agreeableness and Extraversion and (b) Openness and Extraversion. This is in line with recent results in personality psychology that there are general factors of personality that govern factors in the five-factor model [25]. Even though the assumption of independence is violated we can still apply the chosen clustering technique. As pointed out in the instructions of the algorithm [33], “Empirical internal testing indicates that the procedure is fairly robust to violations of both the assumption of independence and the distributional assumptions, but you should try to be aware of how well these assumptions are met.”

Therefore, we decided to proceed with the TSCA in order to explore the grouping structure of our personality data. It is worth mentioning that the TSCA algorithm not only divides the data into clusters, but also includes statistical tests and graphs for the validation of the clusters found.

From the results of the clustering, the AIC criterion determined that the optimal number of clusters was two. Of the 47 participants, 28 were assigned to the first and 19 to the second cluster. Table 2 shows the centroids of the clusters. For each factor a single individual can have a value between 10 and 50 (inclusive). We can see that for all individuals combined the means lie between 33.47 (Extraversion) and 38.43 (Agreeableness).

As can be seen from Table 2, the centroids show that in general the variables separate well in the two clusters. Specifically, participants in Cluster 2 seem to have larger mean in all attributes except conscientiousness where there is little difference between the clusters.

Finally, the algorithm performs a Student’s *t*-test in order to determine which of the variables are really important for the formation of the clusters, in the sense that their mean values are significantly different. The variables are then ranked according to their importance. The result of the test is shown in the bar-plot (corresponding to Cluster 1) of Fig. 1.

The bars represent the *t*-statistic, which is negative when the mean of the variable is lower than the overall mean (as in Cluster 1). The vertical lines indicate two critical values calculated for the *t*-statistic after applying Bonferroni adjustment to each of the cluster numbers [33]. The variables, whose *t*-statistic exceeds these lines is important to the corresponding cluster. In our case, the important variables for both clusters are Extraversion and Openness (in that order of importance). Agreeableness and Neuroticism, although they present differences in the clusters, are not so important while Conscientiousness is not important at all.

Overall we can conclude that we have in our sample two general personality types characterized mainly by differences in Extraversion and Openness, and partly by Agreeableness and Neuroticism. Since Cluster 2 has higher average numerical values

Table 1
Responses to ‘What do you do in your line of work?’ (Respondents could choose multiple alternatives.)

Answer	No. of respondents	%
Programming	40	85.0
Design	25	53.2
Requirements engineering	20	42.5
Customer relations	20	42.5
Documentation	19	40.4
Testing	17	36.2
Verification & Validation	15	31.9
Project management	11	23.4
Requirements elicitation	7	14.9
Personnel management	6	12.8
Other	6	12.8

Table 2
Centroids for the clusters (η = mean, σ = standard deviation) (A = agreeableness, O = Openness, C = Conscientiousness, E = Extraversion, N = Neuroticism).

	A	O	C	E	N
<i>Cluster 1</i>					
η	36.79	34.93	36.29	29.39	35.64
σ	3.446	3.558	3.473	5.513	5.180
<i>Cluster 2</i>					
η	40.84	41.37	36.32	39.47	39.21
σ	5.766	3.700	3.497	5.368	6.079
<i>Combined</i>					
η	38.43	37.53	36.30	33.47	37.09
σ	4.902	4.795	3.445	7.357	5.774

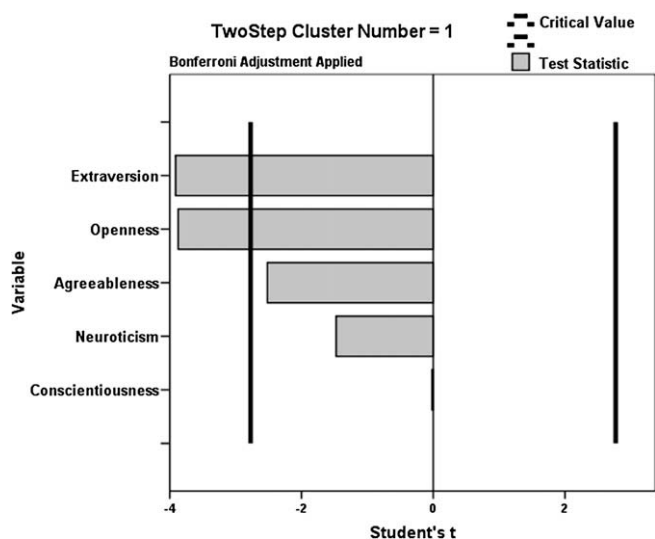


Fig. 1. Bar-plot of the ranked Student's t values for each personality factor for Cluster 1.

for all factors we will sometimes call this cluster the 'intensive' one while the other (Cluster 1) is called the 'moderate' one. We cannot easily portray the five-dimensional space in which these clusters exist, but after a factor analysis we can reduce the dimensions to three, which explain 79% of the overall variance [32]. Then in the three-dimensional space defined by the factor scores, we can plot the respondents. The result is shown in Fig. 2. The clusters are visually distinguishable even in the reduced space. The lines are the distances from their new, three-dimensional centroids.

4.3. Cluster and attitude associations

Given the cluster analysis above and the two clusters found it is now interesting to see how the other variables, having to do with the personal characteristics (e.g. sex or age) and the attitude variables, are distributed with respect to the personality clusters. This

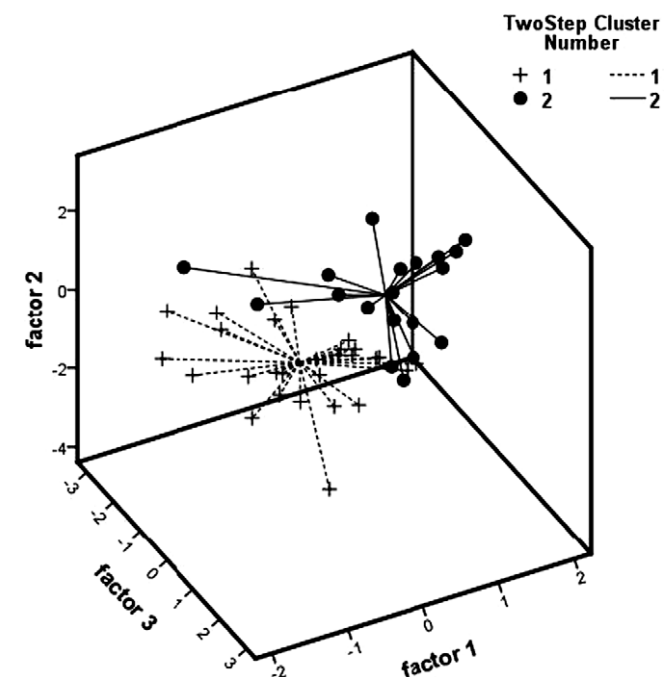


Fig. 2. The two clusters, with distances to their centroids, in a 3D factor space after factor analysis.

will help us to assess the association between them (by using the χ^2 -test of independence). Due to the fact that some variables have many categories and missing values (when respondents did not answer a question the web survey saved a 'No answer' response in the data file), in order to enhance the validity of the χ^2 -test we merge some categories with few observations. In the end, the χ^2 would test if two categorical variables are independent. If this assumption is rejected we can assess dependence (or equivalently association).

Table 3 shows the χ^2 significance levels for a difference between the personality clusters and a subset of the questions from Part II of the questionnaire. They are sorted from highest to lowest significance with levels well above $p = 0.1$ excluded.

The table shows that respondents with a more 'intense' personality (Cluster 2) is more likely to be younger (Q54), prefer doing multiple things at the same time (Q92), prefer contributing to some part of a project instead of working with it from the start to the end (Q96). Female respondents (Q51) were more likely to have a moderate personality (Cluster 1).

There seems to be some dependence for the level of satisfaction with standards the respondents use (Q72) since the majority of respondents who did not claim strong satisfaction belongs to the first cluster while the opposite is true for the satisfied ones. However, due to the small number of respondents that were satisfied, the test gives $p = 0.121$ and, hence, cannot support significant association.

No significant dependence was found for questions in Part II not listed in Table 3.

4.4. ANOVA factor and attitude associations

In order to test the association between each personality factor and all the other categorical variables based on each of the questions in Part II, we performed ONE-WAY ANOVA with the dependent variable being each of the personality attributes separately and as categorical variable the responses to each of the questions. ANOVA shows if there is significant difference in the means of the dependent variable across the categories of the categorical one. When there is such a difference we can assess association between the dependent variable and the factor. Table 4 shows the significance levels between each of the personality factors and a question where a significant ($p < 0.1$) dependence was found. Below we discuss some of the stronger associations (for values of p over 0.05 we write out the actual values) and in Figs. 3–5 some of them are depicted.

Higher levels of agreeableness is associated with preferring to work in teams (question 93), preferring project startup to working on a whole project or doing short contributions throughout a project (96), considering change management important for work satisfaction (83), doing multiple things at once (92), and a higher need to change the current working procedures (74). Agreeableness also seems to be associated with preferring nontechnical tasks above technical tasks (97, $p = 0.075$).

Higher levels of extraversion is associated with performing more efficiently when working according to a schedule or project

Table 3
Significance for difference between clusters.

Question	χ^2 -test on cluster vs.	χ^2 significance level, p
54	Age ^a	0.042
92	Multi-tasking preference	0.080
96	Project responsibility preference	0.082
51	Sex	0.086
72	Satisfaction with standards used	0.121

^a Some of the age categories did not appear in the second cluster. In order to simplify the appearance and perform a valid χ^2 -test we concatenated the original categories into four new ones: 25–30, 31–35, 36–45 and >45.

Table 4
ANOVA significance level for questions and personality factors.

Question	Preference for	<i>E, p</i>	<i>O, p</i>	<i>A, p</i>	<i>N, p</i>	<i>C, p</i>
93	Team work	0.054		0.002		0.099
92	Multi-tasking	0.095	0.045	0.036		
74	Changing work manners			0.046		0.018
83	Change mgmt. importance			0.019	0.089	
88	Current org. reasons for change work				0.098	0.050
97	Non-tech. preference			0.075		0.059
105	Job satisfaction				0.008	
89	Future org. reasons for change work					0.016
96	Project responsibility pref.			0.017		
87	Your reasons for change work					0.019
91	Plan daily work	0.033				
104	Mgmt. supports		0.040			
67	Tool satisfaction					0.045
101	Software engineering research importance					0.086
77	Decisions for higher quality ^a	0.087				
59	Years of experience					0.095
94	Holistic project responsibility		0.099			

^a Categories were merged into two: Low/Quite low and Quite high/High degree.

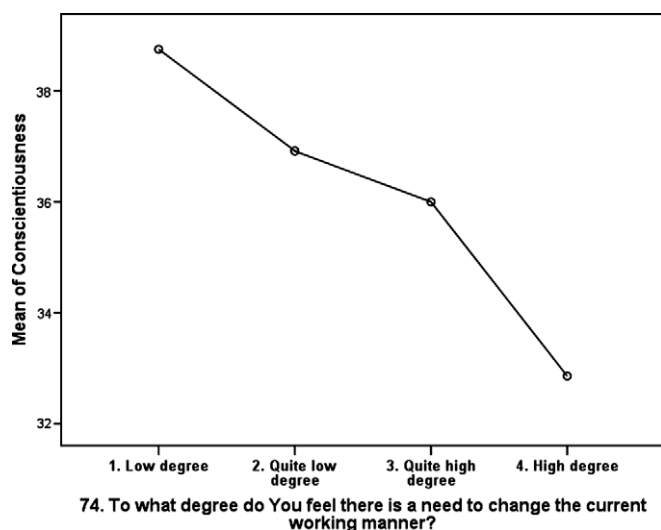


Fig. 3. Mean of conscientiousness vs. question 74.

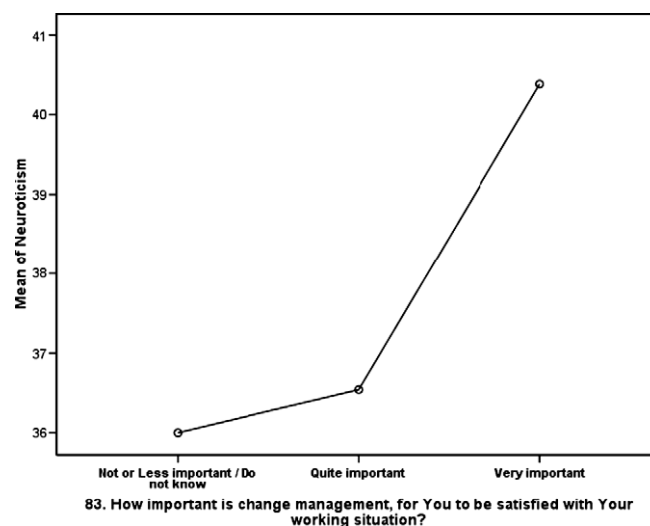


Fig. 5. Mean of neuroticism (higher level indicates more emotionally stable) vs. question 83.

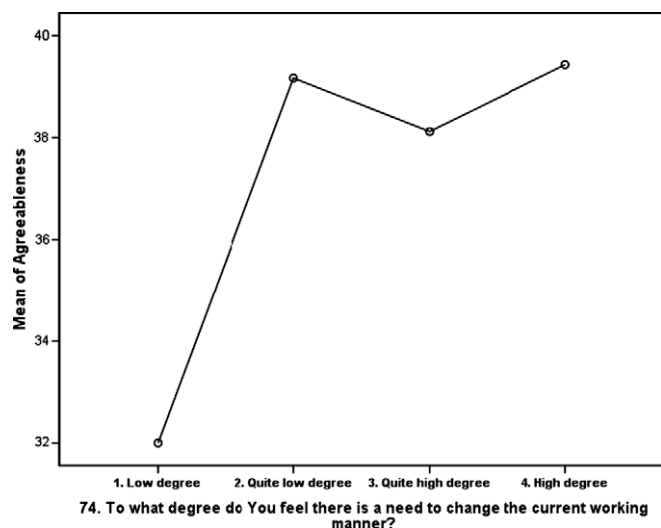


Fig. 4. Mean of agreeableness vs. question 74.

plan (91), and preferring to work in teams (93). Extraversion also seems to be associated with preferring to do multiple things at once (92, $p = 0.095$), and feeling that they can increase the quality of their work to a high or quite high degree by their own decisions (77, $p = 0.087$).

Lower levels of openness is associated with feeling that the management style within the company helps and supports you (104). Higher levels of openness is associated with preferring to do multiple things at once (92). Higher levels also seems to be associated with preferring to take responsibility for a whole project and not single parts (94, $p = 0.099$).

Higher levels of conscientiousness is associated with a low need to change the current working procedures (74). Higher levels also seems to be associated with preferring non-technical tasks above technical tasks (97, $p = 0.059$), thinking software engineering research is not so important (101, $p = 0.086$), and preferring to work alone (93, $p = 0.099$).

Higher levels of neuroticism (i.e. being more emotionally stable) is associated with high or quite high job satisfaction (105). It also seems to be associated with considering change management important for work satisfaction (83, $p = 0.089$).

On questions 59 and 67 the association patterns are complex since the answer alternatives have no clear progression. Questions 87, 88 and 89 are also hard to summarize since there is no progression or scale among the categories in the answer alternatives. The full details can be found in [31].

4.5. Generalized Linear models for factors based on attitudes

In the previous analysis we explored possible associations between personality scores and attitude factors by considering all possible pairs between them. To avoid spurious associations because of multiple comparisons, we here proceed with a multivariate analysis, trying to model the relationship between personality and several attitude variables simultaneously. This involves generating Generalized Linear Models (GLM) in SPSS [36,37,33]. By declaring each one of the personality factors as response variable and the associated attitude questions from Table 4 as predictors, we tried to form a linear equation which expresses the personality score in terms of the estimated effects of the levels of the attitude questions. Below we describe this analysis for each of the personality factors in turn.

For variable E(xtraversion) we found in Table 4 possible association with questions 93, 92, 91 and 77. The GLM estimated from these variables is:

$$E = c + a_{93} + a_{92} + a_{91} + a_{77}$$

where $c = 33.265$ is the intercept,

$$a_{93} = \begin{cases} -3.640 & \text{for answer 'By yourself',} \\ 0 & \text{for answer 'In a team',} \end{cases}$$

$$a_{92} = \begin{cases} -1.118 & \text{for answer 'One thing at a time',} \\ 0 & \text{for answer 'Several things at once',} \end{cases}$$

$$a_{91} = \begin{cases} 4.672 & \text{for answer 'After a given schedule, project plan',} \\ 0 & \text{for answer 'As the day develops',} \end{cases}$$

and

$$a_{77} = \begin{cases} -4.365 & \text{for answer 'Low or Quite low degree',} \\ 0 & \text{for answer 'Quite high or High degree'}. \end{cases}$$

The a -coefficients denote the effects of each attitude factor on the variability of E . All the effects and the intercept were found significant with $p < 0.001$. Also, the whole model was found significant by the likelihood ratio chi-square test used in GLMs with $p < 0.001$, showing that the model explains significant portion of the variability of E when compared to the intercept-only model (i.e. a model that would not take into account the effects of the attitude factors). The sign of the effect is important in order to interpret the positive or negative effect of the factor on the personality variable. Note that the GLM is not intended to be used for prediction but rather to explain the association of each personality variable with several attitude variables simultaneously.

Similarly, for O(penness):

$$O = 33.444 + a_{92} + a_{104} + a_{94}$$

where

$$a_{92} = \begin{cases} -2.675 & \text{for answer 'One thing at a time',} \\ 0 & \text{for answer 'Several things at once',} \end{cases}$$

$$a_{104} = \begin{cases} 3.203 & \text{for answer 'Low or Quite low degree',} \\ 0 & \text{for answer 'Quite high or High degree',} \end{cases}$$

and

$$a_{94} = \begin{cases} 1.998 & \text{for answer 'Entire development process',} \\ 0 & \text{for answer 'Particular part'}. \end{cases}$$

All the effects and the intercept were found significant with $p < 0.001$. Also, the whole model was found significant by the likelihood ratio chi-square with $p < 0.001$.

For A(greeableness):

$$A = 39.551 + a_{93} + a_{92} + a_{74} + a_{83} + a_{97} + a_{96}$$

where

$$a_{93} = \begin{cases} -2.357 & \text{for answer 'By yourself',} \\ 0 & \text{for answer 'In a team',} \end{cases}$$

$$a_{92} = \begin{cases} -0.997 & \text{for answer 'One thing at a time',} \\ 0 & \text{for answer 'Several things at once',} \end{cases}$$

$$a_{74} = \begin{cases} 2.669 & \text{for answer 'Low or Quite low degree',} \\ 0 & \text{for answer 'Quite high or High degree',} \end{cases}$$

$$a_{83} = \begin{cases} -7.372 & \text{for answer 'Not or Less important/Do not know',} \\ 0 & \text{for answer 'Quite or Very important',} \end{cases}$$

$$a_{97} = \begin{cases} 1.877 & \text{for answer 'Soft parts',} \\ 0 & \text{for answer 'Technical parts',} \end{cases} \text{ and}$$

$$a_{96} = \begin{cases} -1.243 & \text{for answer 'From project start to project end',} \\ 0 & \text{for answer 'Project start up or Short contributions'}. \end{cases}$$

All the effects and the intercept were found significant with $p < 0.001$. Also, the whole model was found significant by the likelihood ratio chi-square with $p < 0.001$.

For N(euroticism):

$$N = 38.109 + a_{83} + a_{88} + a_{105}$$

where

$$a_{83} = \begin{cases} -3.796 & \text{for answer 'Not or Less important/Do not know',} \\ 0 & \text{for answer 'Quite or Very important',} \end{cases}$$

$$a_{88} = \begin{cases} 2.373 & \text{for answer 'Customer satisfaction',} \\ -2.759 & \text{for answer 'Enterprising people in the company',} \\ 1.361 & \text{for answer 'Financial reason',} \\ 7.566 & \text{for answer 'Own initiative',} \\ 1.015 & \text{for answer 'Quality of the end product',} \\ 0 & \text{for answer 'Do not know',} \end{cases}$$

and

$$a_{105} = \begin{cases} -5.025 & \text{for answer 'Low or Quite low degree',} \\ 0 & \text{for answer 'Quite high or High degree'}. \end{cases}$$

All the effects and the intercept were found significant with $p < 0.005$ except from the effect of a_{88} for the level 'Quality of the end product' where $P = 0.083$ (meaning that this specific coefficient cannot be proved significantly different from zero although the other levels are significant). Also, the whole model was found significant by the likelihood ratio chi-square with $p < 0.001$.

For C(onscientiousness), the inclusion of all attitude variables from Table 4 in the GLM did not result in a model where all factors contributed significantly. This is reasonable because the number of cases (respondents) is relatively small and because there are various associations among the attitude answers. After removing factors with high p values, from the GLM analysis that SPSS does to show the contribution of each factor to the model, we found the following model:

$$C = 29.895 + a_{74} + a_{89} + a_{87}$$

$$a_{74} = \begin{cases} 1.574 & \text{for answer 'Low or Quite low degree',} \\ 0 & \text{for answer 'Quite high or High degree',} \end{cases}$$

$$a_{89} = \begin{cases} 6.788 & \text{for answer 'Customer satisfaction',} \\ 2.101 & \text{for answer 'Enterprising people in the company',} \\ 3.528 & \text{for answer 'Financial reason',} \\ 5.580 & \text{for answer 'Quality of the end product',} \\ 0 & \text{for answer 'Do not know',} \end{cases}$$

and

$$a_{87} = \begin{cases} -2.689 & \text{for answer 'Customer satisfaction',} \\ 1.003 & \text{for answer 'Enterprising people in the company',} \\ -3.423 & \text{for answer 'Financial reason',} \\ 3.666 & \text{for answer 'Own initiative',} \\ 0.969 & \text{for answer 'Quality of the end product',} \\ 0 & \text{for answer 'Do not know'.} \end{cases}$$

All the effects and the intercept were found significant with $p < 0.005$ except from the effect of a_{89} for 'Enterprising people in the company' where $p = 0.235$, the effects of a_{87} for 'Enterprising people in the company' and 'Quality of the end product'.¹ Also, the whole model was found significant by the likelihood ratio chi-square with $p < 0.001$.

4.6. Intra-question associations

We also analyzed all the Part II questions in pairs with the χ^2 -test. A large number of significant associations were found. Below we just discuss a subset of the most relevant and interesting ones.

Female respondents to a higher degree felt their company encouraged a certain way of working (75), preferred to be responsible for a part of development rather than the whole project (94), and preferred working with non-technical instead of technical parts.

The ability to make decisions that affect the quality of the work is strongly associated with the work being interesting and challenging and also with job satisfaction. Only 11% of the respondents that are very satisfied with their working situation work in an organization with many levels (not a flat organization structure). However, how respondents characterized their organization varied even among respondents in the same organization so it is not clear how valid this question really is.

4.7. Personality dispersion

Since Goldberg explicitly discourage comparisons against norms, very few large-scale norms for the IPIP scales have been presented in the literature [24,27]. However, a study by Buchanan includes results from a five-factor personality test administered over the Internet [38]. Since Buchanan used a different version of the IPIP five-factor scale, which has a different number of items, we cannot directly compare their means. Moreover, we have compared the dispersion in our sample with the one reported on in Buchanan's study.

In our case, we used the coefficient of variation since it is dimensionless and scale-independent and, additionally, can be used to compare populations that have significantly different means [39]. To test significance we used a bootstrap procedure since we did not know the distribution of the coefficient of variation and we had no reason to assume it to be normal [40].

We used 10,000 bootstrap samples for each personality factor and sex (male or female; since the Buchanan data was reported separately for male and female respondents). There was significant difference in dispersion for all factors except for the extraversion factor for the female group. The coefficient of variation was significantly less for our respondents and for all factors except for extraversion among the female respondents. This indicates that the respondents in our study were a more homogeneous group than the respondents in Buchanan's study. Since Buchanan's study involved a larger sample from the general population we have shown that the software engineers in this study are a more homogeneous group, personality-wise, than the general population.

¹ These two alternatives had no significant difference from the 'Do not know' alternative.

5. Discussion

We have divided our discussion into five different parts. First, we discuss threats to validity, followed by the associations uncovered in the empirical study. We then clarify and discuss the method proposed in this paper, how it can be used in practice and ethical concerns in applying it. Part four discusses our experience in using the IPIP personality tests. Finally, we discuss future work.

5.1. Validity threats

A low answer frequency can be a threat to any survey. Since we do not have exact information on how many software engineers were invited by our contact at each company it is hard to evaluate to what degree this is a threat to our results. However, since there are a few companies where only a few engineers answered it seems likely that the answer frequency is generally low even if most of the contact person only emailed the invitation to their closest group. One explanation might be that the study was carried out in the summer; many of the potential respondents might have been on vacation. Another explanation can be that engineers are not used to the use of psychometric tools such as personality tests. With increasing use within the field and in evaluating job applicants this should become less of a problem.

The answer frequencies on questions within the survey were generally high (96–100%) with a few exceptions. We have typically handled this in the analyses by merging categories with few or no answers so this should not be a serious threat.

Since our study is based on self-assessment, and especially since it measures general attributes of the involved respondents, there is a risk for evaluation apprehension. Humans want to 'look good' and 'smart' and it might affect the sincerity of their answers. This is a general problem with personality, and any, tests based on self-assessment. However, the results from several, large empirical studies have shown that the FFM tests such as IPIP are robust to these effects [23]. The same reasoning can be used to address the potential problem that several of the items in the IPIP scales are similar. None of our respondents made any comments about this, but the authors noted it since it might lead to respondents more easily being bored and not being as sincere in their answers. Even though this is less of a problem for the shorter variant of the IPIP scale it should be considered if the longer scale is used.

The total number of respondents is limited so it is a threat to validity that the sample is not representative for a larger group of software engineers. Also there might be national variations that we do not know of. The fact that we only have Swedish respondents working in Swedish companies decreases generalizability. However, large studies on the five-factor personality model have shown no significant differences based on nationality.

5.2. Associations uncovered in the empirical study

Age seems to be strongly associated with decreases in several of the personality factors. One explanation for this is that by having made mistakes and gained experience, people become more moderate in how they approach the world; they lose their youthful naivety. A larger issue is that this can be a serious threat to the idea of personality tests measuring stable characteristics of a person. However, there is nothing in the five-factor model that states that traits are fixed over time even if the five-factor model has been shown to be stable for individuals over decades [23].

The cluster analysis is a powerful statistical method since it can help us make sense of multivariate data. Even though the two clusters are only significantly different on the Extraversion and Openness factors, Agreeableness and Neuroticism also show differences.

In any case, the grouping of respondents into a few groups makes further analyses simpler (less pair-wise comparisons to consider), and is also easier to visualize. This is further simplified by the factor analysis (not to be confused with the personality factors).

Together, cluster analysis and factor analysis are powerful statistical methods that can be used instead of the pre-defined personality types as classified by MBTI and other similar psychometric instruments. Given the more neutral personality testing machinery of IPIP, we can still find different types of personalities by applying these, and potentially other, statistical methods. However, it is more powerful that these types really can be found in the actual data instead of by using pre-defined types that may have little or no support in empirical data. Furthermore, the numerical scales for each factor allows more powerful statistical methods than the dichotomous results of MBTI. Thus by using a more powerful personality instrument together with statistical analysis, as we have shown in this paper, we are less likely to find spurious links and hence more likely to uncover real effects.

Informally we have characterized the clusters as 'intense' vs. 'moderate' personality types. In general, we should avoid such labeling since it might be misleading. We would not like to claim that these are the two main types of software engineering personalities that exist. On the contrary with a larger and more diverse set of respondents, e.g. including non-Swedish, and more and different types of companies and organizations, other or more clusters might be found. Even so it is interesting that there are sub-groups within our group of software engineers. Potentially this could explain the conflicting results that was seen in the systematic review of Beecham et al. as to whether software engineers is a homogeneous group or not [5].

We found four questions in Part II of the questionnaire that associated with the personality-based clusters. Two of them (sex and age) are demographic. While age was discussed above, it is not clear why sex should make a difference when it comes to personality. In contrast with our results, in a previous study female respondents were found to have higher average scores for all personality factors than male respondents [38]. It is not clear to us why sex should associate differently with personality factors among software engineers. It is possible that the software engineering profession would attract different types personalities among women than among men but we have found little previous studies exploring these gender differences. More research, on larger samples, is needed to see if such gender-related links are valid in general.

Even though several associations between personality clusters and attitudes to software engineering was found, for a majority of questions no significant associations were found. A possible explanation might be that the low dispersion, and thus high homogeneity of the sample, makes it harder to find associations. A larger study with a larger sample size would be needed to further investigate this.

The low dispersion relative to Buchanan's larger study is interesting in itself [38]. The analysis of the dispersion, using the coefficient of variance, ensures that it is not the smaller sample size that explains the homogeneity of our sample. Even though it is not unreasonable that people in the same profession share interests and opinions, it is not self-evident that they must also have similar personalities. However, several studies have shown the power of five-factor personality measures to predict vocational interests so it is to be expected [3]. In contrast, Buchanan's larger sample likely includes many different groups and thus should have a larger dispersion [38].

A further advantage with the five-factor model and IPIP tests was seen in the ANOVA analysis, when checking for association between varying levels of each personality factor, and the categorical answers

for each of the questions in Part II. Several interesting associations were found which shows that our method of combining a personality metric like IPIP with detailed statistical analysis can be used to uncover important individual differences that can affect the efficiency of software engineering processes, methods and tools.

We note that there are several strong associations so considering personality issues in software engineering research is called for. Some of the associations we found seem intuitively right. For example, a higher degree of extraversion associates with a higher preference to working in teams. This is intuitive since an extraverted person values social interaction, which is evidently more frequent in a team than when working alone. Other associations are harder to understand. For example, a higher degree of extraversion also associates with performing more efficiently when working according to a schedule or plan. Possibly the latter association can be understood in terms of the first, i.e. since extraverted persons prefer team work, and team work more often requires a schedule or plan (at least in pre-agile development processes), that might be the situations they are more used to.

We note that even if some of the associations we have uncovered can be considered unsurprising, empirical support for them is still needed since there are potentially many such effects that have no basis in reality.

In total we found 14 statistically significant ($p < 0.05$) associations between singular personality factors and individual attitude questions. Another eleven less statistically significant ($0.05 < p < 0.10$) such associations were found. However, it is important to notice that we have analyzed 28 questions and have five factors to consider. Thus a total of 140 possible associations could have been found. So for a large number of factor-question pairs there is no significant association. Thus, as can be expected, it is clear that personality is but one of the possible variables to consider when trying to understand software engineering judgments, decision-making and performance.

We added the statistical analysis using Generalized Linear Models (GLMs) in order to decrease the risk for doubtful associations caused by the inflation of the hypothesis testing error due to multiple comparisons. One advantage of GLMs is that they can be employed when multiple dependent variables are associated, such as in our case when there is a risk that there are dependencies between the answer on different questions in Part II of the questionnaire [36]. The GLM analysis can also show the strength, and not only the existence, of a link between a personality factor and individual answers on attitude questions.

We also note that there is an added benefit in being able to check for associations between personality factors within single personalities and not only between overall personality types as is often the case when using the MBTI.

5.3. Our proposed method

Our empirical study is an example of a general method for establishing empirically based links between people and their software engineering behavior. The proposed method has three main components: one or more psychometric measurement(s), one or more software engineering measurement(s), and statistical methods to analyze the links between the former two. The main idea is that the statistical methods capture knowledge about the links that can later be used to predict software engineering preferences of an individual, and ultimately their behavior and performance, just as it can be generally used to predict job performance [41]. That there is likely to exist links to behavior and performance is supported by general results linking personality factors to team success in product development [42].

The method should be of practical use to managers in software organizations when they form teams and assign project roles. By

being aware of the personalities of their staff and having knowledge of the associations to work preferences they are better equipped to align individual developers to the right role and work tasks. Given the associations we have uncovered, we suggest that software organizations maintain a personality inventory of the staff and that they consider personality factors more widely in their decision making. Not only can this benefit managers and the organization, in getting teams that function better, it can also benefit the employees since they will be more likely to be satisfied with their work. Furthermore, a personality inventory could be consulted when diagnosing problems in ongoing projects, and not only at project setup and planning. However, given the validity threats to our study the manager might need to establish if the associations uncovered in this study are valid in their organization and to what degree. Longer-term more detailed knowledge of preferences would be needed which calls for additional questions to be used when analyzing the associations to personality factors. Software managers and organizations also need to work in cooperation with researchers to establish these more detailed models and knowledge.

Based on our empirical study we suggest that software engineers and researchers choose recent and well-supported psychometric tests that are publicly available. Over time this should increase the usefulness and reusability of the results. One should prefer psychometric tests that have real-valued, continuous outcome variables, since this enables more powerful and nuanced statistical analysis such as the multi-variate analysis we have used here. Preferably, the chosen metric should also have established norms to which the obtained results can be compared. However, this is more important for researchers that are interested in the group of software engineers as a collective. It should be less important for a manager who wants to predict the behavior or performance of his particular team.

For the software engineering measurement this can be chosen at at least three different levels. The most general level is the one that we have used in our study: general attitudes and views about software engineering. A more direct choice would be to measure specific behaviors, such as the occurrence of conflicts between team members in design meetings. The third level would be to measure the performance or outcome of a software engineering activity directly. An example would be the level of customer satisfaction or the resources used in a certain phase. The later and more focused choices have the potential to be more directly useful and important for decision making in software engineering projects. However, the former and less focused choices have the potential to establish more general links that can predict several more specific behaviors and performances. Also the later levels often involve not only individuals but teams of engineers. A limitation of psychometric tests is that they may not be enough to cover such group behavior; sociometric tests might be more appropriate. There is also the potential to study several software engineering components at different levels within the same project or study.

The statistical methods employed will depend on the measures in the other two components. A detailed guide for this choice is outside the scope of this study. However, we have established that a cluster analysis can help in finding overall patterns in the psychometric measurements and also reduce the number of links that needs to be investigated. Cluster analysis will be less useful if the psychometric measurement have few dimensions; for our five dimensions it was useful. We used ANOVA tests to establish connections between each dimension in the psychometric measurements and individual aspects of the software engineering measurement. A more general and powerful technique is to use Generalized Linear Models. This is in line with a general trend that more recent, and often compute-intensive, statistical methods can extract more information from the collected data.

A criticism to the proposed method is that there is no reason to use indirect measures, such as personality, when we can ask developers about their preferences directly. For example, we could ask them directly about their preference for team work. However, an advantage with the general knowledge that the statistical analysis can uncover is that personality can be used in a very large number of 'soft' decisions about team and individuals in software development projects. Essentially, it can act as a substitute for more detailed knowledge about the preferences of the individuals and be a more practically useful tool for a manager or developer. Also directly asking people might not give truthful answers since political and or economical aspects can be weighed in in a particular situation. This threat is less likely when an indirect method is used to establish relations that holds in general.

There are ethical concerns that has to be considered before a study based on this method is carried out. This is often not a problem in studies where only the researchers have access to answers; they can anonymize individual results. However, if a manager leads such a study there is a risk that individual engineers will be reluctant to answer truthfully. Over time this should be less of a problem, in particular since the use of psychometric tests are becoming more widespread. For example, it is common to use personality tests in evaluating applicants when hiring [41]. However, we recommend that companies interested in this type of study teams up with external parties that can guarantee a higher level of anonymity and protection for the results of individual engineers. This is also an excellent opportunity for increased collaboration between software engineering practitioners and researchers.

5.4. Using the IPIP personality tests

In our empirical study, the psychometric measurement is the IPIP 50-item personality scale, which has all of the characteristics discussed in Section 5.3 above. It was easy to download and use the freely available scale and there is little to be lost from administering them over the web. This makes cheap and simple personality testing. In contrast, the more commonly used MBTI is harder to get a copy of and formally requires a licensed assessor to give and interpret the results. Getting such a license is expensive and effectively works as an obstacle to more widespread adoption of personality testing. This is especially so in technical fields, such as software engineering, where researchers cannot be expected have psychological competence or education. In conclusion, the IPIP five-factor tests come with none of these drawbacks. The fact that the 50-item IPIP five-factor scale has only 50 questions compared to the 93 of MBTI can also work to its advantage. When it is combined with other questions or psychometric instruments it is important to keep down the total time needed to complete the survey. Having 50 instead of 93 questions can thus be an advantage, especially when targeting professional software engineers with little extra time to take part in research studies.

On the other hand, the IPIP contains a much larger set of individual scales to choose from and supports many different aggregate tests based on these individual scales. Thus for a researcher, that want to go into more detail and try to find associations at an even more fine-grained level, there are more time-consuming variants that can be employed.

Compared to the MBTI, there are less existing IPIP norms to compare to. In fact we have found only one report with readily available summary statistics [38]. However, the results from that study are based on an older version of the IPIP 50-item scale and thus is not directly comparable to the presently available versions and, in addition, since the five-factor model has been stable over several decades and for several different tests comparisons between old and new versions can be done, at least for indicative purposes.

5.5. Future work

Future studies similar to this one should be conducted to further strengthen the results. In particular we would like to repeat the study but with a larger sample size and in a way so that we can control how many people actually get the invitation to participate. This would allow us to better judge the answer frequency and thus better assess this threat.

There is a potential to investigate the power of other psychometric tests in predicting, explaining and improving software engineering performance [29]. For example there are well-proven temperament tests that can be administered in a similar way to the IPIP [43], e.g. as has been used in software engineering research in [39]. By combining several types of psychometric tests, and potentially involving group-related metrics, potentially better software engineering prediction models can be created. Adding more direct metrics of software engineering performance should also be investigated.

At the most general level, the effects of personality on software engineering judgments and decisions cannot be considered in isolation. As was shown by the Blumberg and Pringle model in Section 2.1 there are many other variables to consider. In particular, the many different variables need to be considered in relation to each other. For example, we want to investigate how motivational factors combine with personality differences, organizational and leadership issues in determining performance in software development teams.

6. Conclusions

This study presented data from an extensive questionnaire taken by 47 software engineers in 10 different software development organizations in Sweden. The questionnaire had two main parts: A 50-item personality test, based on the state-of-the-art five-factor model of personality captured in a public domain test scale, and one part with 56 questions on views, attitudes and procedures in software engineering. Extensive statistical analyses were carried out and reported on in this paper.

A cluster analysis found two clusters of personalities among the respondents: One with more moderate personalities and one with more intense personalities, i.e. scoring higher on the factors of Extraversion and Openness. These clusters showed statistically significant associations with four questions in Part II of the questionnaire; software engineers with more intense personalities were found to be more likely to be younger, to prefer doing multiple things at the same time, and to prefer contributing to some part

of a project instead of working with it from the start to the finish. Female respondents were unlikely to be in this cluster.

A further analysis considered associations between the five personality factors and the responses to the software engineering specific questions in Part II. Several statistically significant associations were found. For example, higher levels of Extraversion is linked to performing more efficiently when working according to a schedule or project plan, and higher levels of Openness is linked to preferring to take responsibility for a whole project and not individual parts. A multi-variate statistical test was also used to quantify the strength of the links between personality factors and the answers on individual questions.

One analysis also considered associations between the questions in Part II. For example, a strong association was found between the ability to take decisions that positively affect the quality of your work and with you perceiving your work as interesting and challenging.

The dispersion in personality within the personality factors were found to be significantly less than for a larger sample of people taking a similar personality test. This indicates that software engineers is a more homogeneous group personality-wise than the population at large. This is in line with similar results for other vocations.

Together, the data and analysis reported on in this paper, finds several strong associations between personality aspects and software engineering related views, attitudes and work preferences. Given that state-of-the-art personality tests with good support in the psychological theory and literature are nowadays readily available on the Internet, we believe that more software engineering researchers and practitioners should add personality metrics to their set of tools. In combination with powerful statistical techniques readily available in software packages they form an excellent method and tool set to assess the power of personality on software engineering judgments and decisions, and thus ultimately on performance. This knowledge can be used by managers to more effectively form teams and analyze problematic projects. The method we propose can thus be an important step towards evidence-based software engineering that considers human factors.

Acknowledgments

The authors want to thank all the 47 participants in the 10 companies that took part in this research and their managers for allowing them to participate. We also want to acknowledge the helpful comments from the anonymous reviewers. This research was supported partly by the Knowledge Foundation in Sweden (under a research grant for the project BESQ).

Appendix A

A.1. Questionnaire Part II

Part I of the Questionnaire was the 50-item IPIP scale. Thus, questions in Part II was numbered from 51 and onwards. Tables 5–11 summarize the questions that was used in the statistical analysis reported on in this paper. However, we have also included questions (64, 65, 69, 70) that are needed to understand the ana-

Table 5
Questionnaire Part II – you and your tasks.

No.	Question	Answer alternatives
51	Gender?	Male, female
54	Age?	<25, 25–30, 31–35, 36–40, 41–45, 46–50, 51–55, 56–60, >60
59	Years of experience in software development?	<2, 2–5, 6–10, >10
63	Typical size of projects in which you work?	<3 months, 3–6 months, 7–12 months, >12 months

Table 6
Questionnaire Part II – tools.

No.	Question	Answer alternatives
64*	Which tools do you use daily?	ArgoUML, Eclipse, Parasoft, Rational Rose, SunOne, TogetherSoft, Visual Studio, WithClass, DeSign, Other
65*	Which tool is most important in your work?	Same as for Q. 64
67	How satisfied are you with that tool?	Very satisfied, Somewhat satisfied, Somewhat dissatisfied, Not satisfied, Do not know

Table 7
Questionnaire Part II – standards.

No.	Question	Answer alternatives
69*	Which standards affects your daily work?	CMM (Capability Maturity Model), ISO 12207 (Software Life Cycle Process), ISO 15288 (Life Cycle Management), ISO 9000 (Quality Management), P-CMM (People Capability Maturity Model), PSP (Personal Software Process), RUP (Rational Unified Process), SPICE (Software Process Improvement and Capability dEtermination), TSP (Team Software Process), <i>Other</i>
70*	Which standard do you consider most important in your work?	<i>Same alternatives as in Q. 69 above</i>
72	How satisfied are you with that standard?	Very satisfied, Somewhat satisfied, Somewhat dissatisfied, Not satisfied, Do not know

Table 8
Questionnaire Part II – about your organisation.

No.	Question	Answer alternatives
74	To what degree do you think there is a need to change the working methods you use today?	High degree, Quite high degree, Quite low degree, Low degree, Do not know
75	To what degree do you think your company encourages the use of specific working methods?	<i>Same alternatives as in Q. 74 above</i>
76	To what degree do you think you can decide on the working methods to use?	<i>Same alternatives as in Q. 74 above</i>
77	To what degree do you think you can take decisions that affect the quality of your work?	<i>Same alternatives as in Q. 74 above</i>
79	To what degree do you think your work is interesting and challenging?	<i>Same alternatives as in Q. 74 above</i>
83	How important is improvements/change management for you to be satisfied with your working situation?	Very important, Quite important, Less important, Not important, Do not know
87	What is Your primary reason for changing the working methods?	Enterprising people in the company, Own initiative, Financial reasons, Adhering to industry standards, Customer Satisfaction, Quality of the end product, Do not know, <i>Other</i>
88	What do you think is the company's primary reason for changing the working methods?	<i>Same alternatives as in Q. 87 above</i>
89	What do you think should be the company's primary reason for changing the working methods?	<i>Same alternatives as in Q. 87 above</i>

Table 9
Questionnaire Part II – your preferences.

No.	Question	Answer alternatives
91	Do you prefer to work:	After a given schedule/project plan, As the day develops
92	Do you prefer working with:	Several things at once, One thing at a time
93	Do you prefer working:	In a team, By yourself
94	Do you prefer to be responsible for:	Entire development process, Particular part
95	Do you prefer to work in:	Longer projects lasting for several months up to a year, Short projects lasting up to a couple of months
96	Do you prefer to work:	On project startup, From project start to project end, Short contributions as needed
97	If you could choose would you prefer to work with:	Technical parts, 'Soft' parts
98	You work best/most efficiently when:	Your manager prioritizes Your tasks, You can prioritize Your own tasks

Table 10
Questionnaire Part II – software engineering research.

No.	Question	Answer alternatives
100	How important is it with research towards new working manner in software engineering?	Very important, Quite important, Less important, Not important at all, Do not know

Table 11
Questionnaire Part II – company profile.

No.	Question	Answer alternatives
101	Which of the following alternatives is the closest match to Your company?	Many levels with active management from Your closest superior and almost no cooperation between departments, Relative flat organisation where the closest manager leads the day to day business, Closest manager works as a part of the team. Great possibility to influence as an employee
102	To what degree do You feel that the management style within the company gives you opportunity to influence?	High degree, Quite high degree, Quite low degree, Low degree, Do not know
104	To what degree do You feel that the management style within the company supports and helps you in your work?	High degree, Quite high degree, Quite low degree, Low degree, Do not know
105	How satisfied are You with Your working situation?	High degree, Quite high degree, Quite low degree, Low degree, Do not know

lyzed questions; these are marked with a star next to their number. The original questions were in Swedish and have been translated for inclusion in the paper. The questions that are not included here are available from the authors upon request. In the tables, an answer alternative *Other* indicates that there was an open text area where the respondent could type in any answer.

References

- [1] D. Broadbent, *Decisions and Stress*, Academic Press, 1971.
- [2] G. Matthews, D.R. Davies, S.J. Westerman, R.B. Stammers, *Human Performance: Cognition, Stress and Individual Differences*, Psychology Press, 2000.
- [3] Jesus Salgado, The five factor model of personality and job performance in the European community, *Journal of Applied Psychology* 82 (1) (1997) 30–43.
- [4] A. Furnham, K. Petrides, C.J. Jackson, T. Cotter, Do personality factors predict job satisfaction?, *Personality and Individual Differences* 33 (8) (2002) 1325–1342.
- [5] S. Beecham, N. Baddoo, T. Hall, H. Robinson, H. Sharp, Motivation in software engineering: a systematic literature review, *Information and Software Technology* 50 (9–10) (2007) 860–878.
- [6] G.M. Weinberg, *The psychology of computer programming*, van Nostrand Reinhold Company, 1971.
- [7] B. Curtis, A review of human factors research on programming languages and specifications, in: *Proceedings of the 1982 Conference on Human Factors in Computing Systems*, ACM Press, New York, NY, USA, 1982, pp. 212–218.
- [8] T. DeMarco, T. Lister, *Peopleware: Productive Projects and Teams*, Dorset House Publishing Co., 1987.
- [9] M.I. Kellner, B.C.T. DeMarco, K. Kishida, M. Schlumberger, C. Tully, Non-technological issues in software engineering, in: *ICSE'91: Proceedings of the 13th International Conference on Software Engineering*, IEEE Computer Society/ACM Press, 1991, pp. 144–146.
- [10] A. Howard, Software engineering project management, *Communications of the ACM* 44 (5) (2001) 23–24.
- [11] L.F. Capretz, Personality types in software engineering, *International Journal of Human–Computer Studies* 58 (2) (2003) 207–214. <[http://dx.doi.org/10.1016/S1071-5819\(02\)00137-4](http://dx.doi.org/10.1016/S1071-5819(02)00137-4)>.
- [12] J. Chao, G. Atli, Critical personality traits in successful pair programming, in: *AGILE*, IEEE Computer Society, 2006, pp. 89–93.
- [13] J. Karn, T. Cowling, A follow up study of the effect of personality on the performance of software engineering teams, in: *ISESE'06: Proceedings of the 2006 ACM/IEEE International Symposium on Empirical Software Engineering*, ACM Press, New York, NY, USA, 2006, pp. 232–241.
- [14] D.C. Smith, The personality of the systems analyst: an investigation, *ACM SIGCPR Computer Personnel* 12 (2) (1989) 12–14.
- [15] A.D.D. Cunha, D. Greathead, Does personality matter? An analysis of code-review ability, *Communications of the ACM* 50 (5) (2007) 109–112. <<http://doi.acm.org/10.1145/1230819.1241672>>.
- [16] S. Acuna, N. Juristo, Assigning people to roles in software projects, *Software – Practice and Experience* (34) (2004) 675–696.
- [17] M. Blumberg, C.D. Pringle, The missing opportunity in organisational research: some implications for a theory of work performance, *Academy of Management Review* 7 (4) (1982) 560–569.
- [18] R.L. Atkinson, R.C. Atkinson, E.E. Smith, D.J. Bem, *Introduction to Psychology*, 11th ed., Harcourt Brace Jovanovich, FL, USA, 1993.
- [19] J. Sabini, *Social Psychology*, second ed., W.W. Norton & Co Inc., 1995.
- [20] C.G. Jung, *Psychological Types*, *Collected Works of C.G. Jung*, vol. 6, Princeton University Press, 1971.
- [21] I.B. Myers, M.H. McCaulley, N.L. Quenk, A.L. Hammer, *MBTI Manual (A Guide to the Development and Use of the Myers Briggs Type Indicator)*, third ed., Consulting Psychologists Press, 1998.
- [22] R.R. McCrae, P.T. Costa Jr., Reinterpreting the Myers–Briggs type indicator from the perspective of the five-factor model of personality, *Journal of Personality* 57 (1) (1989) 17–41.
- [23] R.M. McCrae, O.P. John, An introduction to the five-factor model and its applications, *Journal of Personality* 60 (2) (1992) 175–215.
- [24] L.R. Goldberg, A. Broad-Bandwidth, Public-domain, personality inventory measuring the lower-level facets of several five-factor models, in: *Personality Psychology in Europe, Selected Papers from the Eighth European Conference on Personality held in Ghent, Belgium, July 1996*, vol. 7, 1999, pp. 7–28.
- [25] J.P. Rushton, P. Irwing, A general factor of personality (GFP) from two meta-analyses of the big five: Digman (1997) and , Mount, Barrick, Scullen, and Rounds (2005), *Personality and Individual Differences* 45 (7) (2008) 679–683.
- [26] P.T.C. Jr., R.R. McCrae, *NEO PI-R Professional Manual*, Psychological Assessment Resources, Inc., 1992.
- [27] L.R. Goldberg, J.A. Johnson, H.W. Eber, The international personality item pool and the future of public-domain personality measures, *Journal of Research in Personality* 40 (1) (2006) 84–97.
- [28] J.P. Rushton, P. Irwing, The Myers–Briggs personality type and its relationship to computer programming, *Journal of Research on Computing in Education* 26 (3).
- [29] R. Feldt, R. Torkar, L. Angelis, M. Samuelsson, Towards individualized software engineering: empirical studies should collect psychometrics, in: *ICSE Workshop on Cognitive and Human Aspects on Software Engineering (CHASE'08)*, IEEE, 2008.
- [30] R. Likert, A technique for the measurement of attitudes, *Archives of Psychology* 23 (140) (1932) 1–55.
- [31] M. Samuelsson, *Personality Types and Attributes in Software Engineering*, Master Thesis, Dept. of Informatics, University West, Trollhättan, Sweden, June 2005.
- [32] D.J. Bartholomew, F. Steele, I. Moustaki, J.I. Galbraith, *The Analysis and Interpretation of Multivariate Data for Social Scientists (Texts in Statistical Science)*, Chapman & Hall/CRC, 2002.
- [33] SPSS Inc., *SPSS for Windows*, <<http://www.spss.com>> (last visited: 20.03.07).
- [34] H. Akaike, A new look at the statistical model identification, *IEEE Transactions on Automatic Control* 19 (6) (1974) 716–723.
- [35] D.E. Knuth, *The Art of Computer Programming: Seminumerical Algorithms*, third ed., Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1997, vol. 2.
- [36] A. Dobson, A. Barnett, *Introduction to Generalized Linear Models*, third ed., Chapman and Hall/CRC, 2008.
- [37] P. McCullagh, J. Nelder, *Generalized Linear Models*, Chapman and Hall, 1989.
- [38] T. Buchanan, *Online Implementation of an IPIP Five Factor Personality Inventory*, March 2007. <<http://www2.wmin.ac.uk/buchant/wwwffi/introduction.html>>.
- [39] P. Sfetos, I. Stamelos, L. Angelis, I.S. Deligiannis, Investigating the impact of personality types on communication and collaboration-viability in pair programming—an empirical study, in: P. Abrahamsson, M. Marchesi, G. Succi (Eds.), *XP, Lecture Notes in Computer Science*, vol. 4044, Springer, 2006, pp. 43–52.
- [40] B. Efron, R.J. Tibshirani, *An introduction to the bootstrap*, *Monographs on Statistics and Applied Probability*, vol. 57, Chapman and Hall, 1993.
- [41] M.G. Rothstein, R.D. Goffin, The use of personality measures in personnel selection: what does current research support?, *Human Resource Management Review* 16 (2) (2006) 155–180, doi:10.1016/j.hrmr.2006.03.004.
- [42] S.L. Kichuk, W.H. Wiesner, The big five personality factors and team performance: implications for selecting successful product design teams, *Journal of Engineering and Technology Management* 14 (3–4) (1997) 195–221, doi:10.1016/S0923-4748(97)00010-6.
- [43] D. Keirse, M. Bates, *Please Understand Me: Character and Temperament Types*, third ed., Prometheus Nemesis Book Company, 1984.