# An Exploration of Individual Personality Types in Software Development

Murat Yilmaz[1,3] and Rory V. O'Connor[2,3] and Paul Clarke[4,3]

[1] Çankaya University, Turkey
myilmaz@cankaya.edu.tr
[2] Dublin City University, Ireland
[3] Lero, the Irish Software Engineering Research Center
roconnor@computing.dcu.ie
[4] Regulated Software Research Centre, Dundalk Institute of Technology, Ireland
paul.clarke@dkit.ie

**Abstract.** Previous research - using conventional psychometric questionnaires - has highlighted the importance of aligning compatible personality types in software development teams. However, there does not exist a dedicated, robust questionnaire instrument for revealing the pertinent personality types for software development practitioners. This study analyzes the validity and reliability of a 70-item (context dependent) personality-profiling questionnaire particularly developed to assess personality types of software practitioners. A systematic process of validation, using an iterative approach to questionnaire development, was employed. The questions were developed both with a qualitative analysis of interview data, and based on the opinions of expert reviewers who revised the items through a set of examination. To investigate how stable the questions and reproducible the results, we measured test-retest reliability of the instrument, yielding satisfactory results. The present study provided evidence for the construct validity of the instrument. Ultimately, an initial comparison of the results delivered by the instrument demonstrated positive correlations with the findings acquired with well-known personality assessment instrument, i.e. the big five personality questionnaire.

**Keywords:** Software Developers Personalities, Personality Profiling, Myers-Briggs Type Indicator (MBTI) , Questionnaire Validation.

## 1 Introduction

Software projects face several challenges in their dynamically changing organizational environments [1]. These challenges form perceived productivity differences among software practitioners who have a number of distinctive personality types. Considering software development as a socio-technical practice, members of a software team should interact and follow a software development process [2]. One of the key components of success in a software development organization is

selection of the right employee or a team for the right tasks [3]. Indeed, compatibility of practitioners' personalities becomes an important concern for the team success [4]. It is therefore not surprising to discover that several researchers in the field of software engineering have focused on the effects of personality types on the software development process and organizational performance [5–7].

In today's software engineering landscapes, technical skills of the individuals should certainly match with the required talents and experience. In addition, to place the individuals in the right groups or jobs, the social aspects such as individuals' compatibility within a team has emerged as a research interest. This requires a new way of understanding the personality differences with a focus on personality types over structural configurations [8]. The notion of MBTI classifies personality types via four dichotomous dimensions; extroversion-introversion (E-I), which shows the methods for an individual to draw energy (outer word versus inner word of ideas), sensing-intuitive (S-N), which refers the methods of individuals to process data (facts versus possibilities), thinking-feeling (T-F), which is related with people's decision making (objective versus subjective), and judging-perceiving (J-P) identifies whether an individual has a structural or an adaptable style to deal with the word. Although the Myers-Briggs Type Indicator (MBTI) is the most well known and widely used self report instrument in the software industry, some critics argue that individuals may have problems in distinguishing their true preferences from socially desirable type of responses. In fact, some researchers hold the view that the MBTI test results may not reflect the true personality types of individuals [9, 10]. As it is a self-report measure, thus, some may fake their answers. To deal with such a problem, we envision that such a test should be conducted as an interactive face-to-face assessment where the context of the questions on such a dichotomous personality scale should be based on real situations and events extracted from software development industry. In support, Kaluzniacky argues that a personality instrumental specific to IT should be based on a IT-related content while being parallel with the original MBTI [11].

In light of these remarks, we propose a personality-profiling questionnaire, which was tested on practitioners both from academia and industry. The purpose of the present study is to investigate the developed personality-profiling tool for software practitioners and explore its reliability and validity. To evaluate its validity and internal consistency, we carried out a validation process where we assessed the aspects of content validity, and performed a factor analysis for the hypothesized 4-factor personality model. To compare the results of the assessment with other instruments, we conducted well known and a comprehensive psychometric questionnaire (i.e. Big Five questionnaire [12]) on a selected group of participants. Based on a five-factor model, the big five personality inventory is a measurement device, which is used to assess the personality traits.

## 2 Method

This study was conducted in six main steps. Although, it was conducted with both academia and industry in different steps of the work, the main part of the was performed within a middle-sized software company. During all industrial assessments, interviewees were selected by the managerial team of the software development organization. In the assessments conducted in academia, we used individuals, who were novice developers with at least a year of industrial experience. These individuals were picked by the criterion of either whether either they have worked together as a team for some projects or they are the individuals who worked in the same environment at least for sometime.

An illustration of all the steps used in the present study is shown in Figure 1
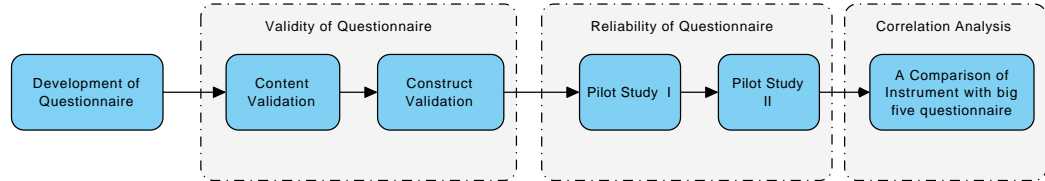


**Fig. 1.** The steps of the study.

The first step is the question development process where a 70-item questionnaire was constructed based on a number of interviews with twenty software practitioners and selected experts from the company. In the second step, we critically reviewed the content of the questionnaire with a group of experts from the managerial team of the company to validate its content. For the third step, we investigated the assessment validity, which is based on the analysis of the correlations among the proposed questions using confirmatory factor analysis with 213 participants from the middle-sized software company. In the fourth and fifth step, to assess the reliability, a pilot study was conducted with a group of novice developers twice (six weeks after the first round of data collection) in a university environment.

Lastly, using 15 participants we conducted an alternative big five personality assessment, and the two assessment scores were compared using correlation analysis. The descriptions of all those involved are shown in Table 1.

### 2.1 Development of Questionnaire

In the first part of our approach, we carefully investigated the 70 questions of Keirsey's temperament sorter [13], which encompasses all four personality types of the MBTI scale as an initial template. Using 10 academics from a psychology

| Step | Number of participants | Type of Assessment |
|------|------------------------|--------------------|
| 1st | 20 (12 males, 8 females) | Interviews |
| 2nd | 10 (8 Males, 2 females) | Experts Reviewers |
| 3rd | 213 (165 Male, 48 Female) | Questionnaire Validation |
| 4th | 15 (9 males, 6 females) | Reliability Study I |
| 5th | 15 (9 males, 6 females) | Reliability Study II |
| 6th | 15 (9 males, 6 females) | Comparison Study |

**Table 1.** Research steps and number of participants

department and 10 highly experienced software practitioners, we assessed the personality types of participants with Keirsey's sorter. Next, we conducted a total number of 30 semi structured interviews where we first asked 5 selected individuals a number of questions that were asked verbally such as *"What do you think about a content specific personality type assessment?"*, *"What kind of events or situations have you observed in software development landscapes that can be useful for an new kind of assessment?"*. Using the acquired information, we built a new set of questions and asked 10 other participants *"What about this situation; do you think about it may be able to reflect the characteristics of a software practitioner?"*, *"How would you react such a question?"*. After performing a rigorous analysis of collected data, we formed 70 questions regarding software engineering context and conducted 15 more follow-up discussions with the participants, and asked them each question *"How do you find the tone of such a question?"*, *"Do you think that question sound right?"*. Lastly, the final form of the personality-profiling questionnaire was discharged to 6 participants who had previously completed the earlier version of the questionnaire - and the results of both engagements were compared.

Table 2 outlines the profile of the 20 participants including their roles (titles), age, years of experience and level of education.

## 2.2 Content Validation

As a second step, we worked directly with 10 experts for validating the content of the questionnaire. The selection of expert reviewers was mainly based on their experience on the field and scientific qualifications. The candidates were proposed to the middle-sized software company and the selection process was executed by the management team. At this point, we conducted several expert review sessions and two panel discussions where all questions were investigated independently by experts with view to identifying problematic items. The participants were required to rate the content for clarity, readability, relevance, etc. on a 4-point Likert scale (1=not clear, 4=very clear).

| Title | Age | Years of Experience | Education |
|---|---|---|---|
| IT Specialist | 33 | 6 | MSc. |
| Project Manager | 47 | 7 | PhD. |
| Software Architect | 37 | 12 | BSc. |
| Software Developer | 31 | 6 | BSc. |
| Software Developer | 33 | 7 | BSc. |
| R&D Team Lead | 39 | 14 | PhD.. |
| Software Tester | 32 | 4 | MA. |
| System Analysis | 34 | 9 | BA. |
| R&D Team Member | 32 | 7 | MSc. |
| R&D Team Member | 31 | 5 | MSc. |
| Organizational Psychologist | 49 | 17 | PhD.. |
| Clinical Psychologist | 57 | 20 | PhD. |
| Student Psychologist | 24 | 1 | BA. |
| Student Psychologist | 22 | 1 | BA. |
| Novice Psychologist | 26 | 2 | BA. |
| Novice Psychologist | 26 | 2 | BA. |
| Student Psychologist | 24 | 1 | BA. |
| Student Psychologist | 23 | 1 | BA. |
| Student Psychologist | 25 | 1 | BA. |
| Industrial Psychologist | 54 | 14 | MA. |

**Table 2.** Participants' Information

Table 3 shows the profile of the 10 expert reviewers including their roles (job titles), age, years of experience and level of education.

| Expert ID | Title | Age | Years of Experience | Education |
|---|---|---|---|---|
| E1 | Software Manager | 46 | 20 | PhD. |
| E2 | UX Designer | 36 | 7 | MSc. |
| E3 | Graphical Designer | 30 | 4 | BA. |
| E4 | Software Practitioner | 31 | 6 | BSc. |
| E5 | Clinical Psychologist | 43 | 16 | PhD. |
| E6 | Organizational Psychologist | 39 | 11 | PhD. |
| E7 | Instructional Designer | 38 | 9 | MA. |
| E8 | Assist. Prof. Dr. (Researcher) | 40 | 14 | PhD. |
| E9 | Assoc. Prof Dr. (Researcher) | 45 | 17 | PhD. |
| E10 | Prof. Dr. (Researcher) | 58 | 25 | PhD. |

**Table 3.** Expert Reviewers' Information

Next, we quantified the extent of agreement among the participants. Based on the experts' ratings, the content validity index (CVI) was calculated as follows:

$$CVI = \frac{number\ of\ raters\ giving\ a\ rating\ of\ 3\ or\ 4}{Total\ number\ of\ raters} \tag{1}$$

where CVI is a coefficient when calculated as 0 indicates that there is a total lack of agreement among participants, and a value 1 shows a total agreement among the experts [14]. After having a discussion with the experts, we had an agreement that items should be accepted when CVI is higher than 80%. A value between 70% and 79% was considered questionable whereas all items below that threshold were considered as unacceptable for validity of its content. Among the 70 questions, 48 questions were rated as accepted. In addition, 12 questions were

found questionable (between 70% and 79%) where 8 items were found below the threshold. All questionable items were revised and all unacceptable items were completely changed regarding the reviews and further sent back to experts for rating. Finally, the finished questionnaire was discussed with experts from the software engineering field.

## 2.3   Assessing Construct Validity

This part of the paper investigates the construct validity with respect to personality characteristics similar to Myers-Brigs approach where four dichotomies were assumed to be the identifiable by the questions, which were asked during the assessment. To assess the validity of the developed questionnaire, the test scores were interpreted with respect to the understanding of participants and the researcher. To analyze the correlation among the questions of the instrument, personality assessment was conducted on 213 software practitioners (Cronbachs alpha was .86). We used factor analysis method where we identified four clusters of questions compatible with the personality constructs all of which were highly correlated. For example, a question regarding social interactions was found to be correlated with extroversion-introversion.

To investigate the hypothesized four factors of personality using the questionnaire, LISREL [15] is used to perform confirmatory factor analysis of the measurement items, and test the five factor model. The hypothesized model had statistically significant values of all of the factor loadings that were between .51 and .80 ($p < .05$). The independence model was clearly rejectable where the $\chi^2$ for independence model with 170 degrees of freedom was 922.319. The proposed model yielded a good-fit[1], where $\chi^2(186, N = 192) = 242.505, p < 0.001$, and the fit indices for the proposed model were satisfactory; $RMSEA = .0615, GFI = .87, AGFI = .82, CFI = .896, NNFI = .88$). Furthermore, a $\chi^2$ difference test was conducted, $\Delta\chi^2(24, N = 192) = 679.814, p < 0.001$).

Overall, the hypothesized factor-based model for personality, which was proposed based on the questionnaire items: EI (10 items), SN (20 items), TF (20 items) and JP (20 items) were assessed using empirical data.

Table 4 illustrates the number of items with highest loading values for the four factor design extracted in this study. For this model, loading on the first factor was the EI scale, which has all of its items. Although some items were found to be loading a few items in other scales, the offending items were essentially evenly distributed on the other scales. The four factor model approach seems to fit with the data where questions were found to be correlated with the constructs that they were suppose to measure. Overall, the analysis suggested that the structure of the questionnaire showed evidence of satisfactory item-to-scale structure.

---

[1] RMSEA = root mean square error of approximation; NNFI = nonnormed fit index; CFI = comparative fit index; AIC = Akaike information criterion.

| | Factor 1 | Factor 2 | Factor 3 | Factor 4 |
|---|---|---|---|---|
| EI | 10 | 0 | 0 | 0 |
| SN | 3 | 17 | 0 | 0 |
| TF | 3 | 1 | 16 | 0 |
| JP | 0 | 0 | 2 | 18 |

**Table 4.** Number of Factors with the highest values of loading in 4-Factor Design

### 2.4 Assessing Reliability

To investigate the reliability of the questionnaire, we assessed 15 novice software practitioners who were studying at a university. The participants who were selected had to satisfy the following criteria: They have all either worked as a team member previously or collaborated in a software project. Biemer suggests that the test-retest method for a questionnaire is one of the most common methods for investigation especially to identify errors of measurement [16]. In support, Presser et al. [17] indicate that parallel measurements conducted to replicated the original assessment is a useful method for assessing the quality of a survey, which can highlight problematic questions.

To assess the reliability of the instrument, the developed questionnaire was conducted twice with the same participants over a six weeks period. Without changing any environmental conditions, the reliability of responses was investigated with the replication study where same questions were asked to the same set of participants. At this point, the goal of the study was to observe "*the measurement error variance associated with the original survey response*" [16, pp. 298]. Using such a re-measurement approach, we analyzed the response of the participants for each question in the questionnaire to calculate "the ratio of question-level measure of response variance to the total response variances" for a given question termed as *index of inconsistency* ($I$) [18] whereas the reliability ratio ($1 - I = \kappa$) is also known as Cohen's measurement of reliability [19]. ($I$) can be represented as

$$I = \frac{g}{p_1 q_2 + p_2 q_1} \tag{2}$$

where total sample size is $n = a + b + c + d$ where $a$ indicates the individuals who select the first option in both interviews, $d$ shows the participants who select the second option in both runs. The number of participants who chose the first option in the first run and second option in the second run is denoted as $c$. Lastly, b is the number of individuals who select the second option in the first run and first option in the second run. Furthermore, $g = (b + c)/n$ is considered as the rate of disagreement, the ratio $p_1 = (a + c)/n$ shows the first answer in the first run, where for the second run the ratio is shown as $p_2 = (a + b)/n$.

To check the reliability of each question, $\kappa$ values were obtained where $\kappa$ ranges are shown in Table 5. During the analysis, we observed that these values are too sensitive, i.e. even one different answer could change the course of results very significantly. After discussing the sensitivity with an expert, we decided to

chose 30% as a cut-off range for the question, and therefore the questions Q4, Q21, Q22, Q24, Q26, Q27, and Q31 were found below the expected value.

| $\kappa$ % Range | Number of Questions |
|---|---|
| 0 - .30 | 7 |
| .31 - .45 | 9 |
| .46 - .60 | 10 |
| .61 - .75 | 14 |
| .76 - .90 | 30 |

**Table 5.** The Range of $\kappa$ numbers found for the academic pilot study

Further analysis showed that one question type from extroversion-introversion, and two question from other types were found to unreliable. By revisiting several experts once again, we discussed the possible updates and alterations for the identified questions. Later, the questions were readjusted. Additionally, we conducted the assessment on 15 software practitioners on a middle-sized software company. Once again, the test was replicated after six weeks to measure the reliability of each question. The results obtained from the analysis of $\kappa$ values for the questionnaire are presented in Table 6.

| $\kappa$ % Range | Number of Questions |
|---|---|
| 0 - .30 | 1 |
| .31 - .45 | 9 |
| .46 - .60 | 12 |
| .61 - .75 | 18 |
| .76 - .90 | 30 |

**Table 6.** The Range of $\kappa$ numbers found for the pilot study

Data from Table 5 can be compared with the data in Table 6, which shows that two questions from 30% range has moved to 60% range, whereas four questions were moved to 75% range, hence we were able to improve the reliability of six more questions after having various iterations for fine-tuning the questionnaire.

From this data, we can see that this part of the study yielded relevant results where only 7 of 70 questions were found problematic (one question from (E/I) trait, and two questions from each (S/N), (T/F), (J/P) traits were out of range). Therefore, we performed our calculations by dropping these questions, and the ultimate results are shown in Table 7.

What is interesting in this analysis is that extroversion was observed as a dominant dichotomy during the pilot study, which is compatible with the recent findings in MBTI research in the field of software engineering (see e.g. [20]).

To reinterpret our work in terms of the five factor model, using a randomly selected subset of software practitioners, we correlated the results of the assessment device and the Big Five personality constructs. To this end, among the participants, 15 of them were assessed with the Big Five Inventory (BFI-44 item

| MBTI Type | Number of Participants | % in Sample Population |
|---|---|---|
| ENFJ | 2 | 13 |
| ENTJ | 2 | 13 |
| ESFJ | 3 | 20 |
| ESFP | 3 | 20 |
| ESTP | 1 | 7 |
| INTP | 1 | 7 |
| ISFJ | 2 | 13 |
| ISFP | 1 | 7 |
| Total | 15 | 100 |

**Table 7.** Personality Types found in the Pilot Study

scale) [12]. Next, the results acquired from the two scales were compared using correlation analysis (see Table 8).

Correlations between the results of the proposed scale, and BFI scale are shown in Table 8. It can be seen from the table that EI was highly correlated with the extroversion, and SN scale was correlated with openness. There were also significant relations between SN versus agreeableness, and JP versus conscientiousness. However, there were no significant relationships were found between any subscales with neuroticism ($N = 15$, $p < 0.01$). The findings of the current study are consistent with those of McCrae and Costa who observed significant correlations between MBTI scale and big five personality characteristics [21].

| | Extraversion | Openness | Agreeableness | Conscientiousness | Neuroticism |
|---|---|---|---|---|---|
| EI | **0.82** | 0.02 | -0.01 | -0.15 | 0.10 |
| SN | 0.11 | **0.58** | 0.03 | -0.08 | 0.01 |
| TF | 0.17 | 0.02 | **0.49** | -0.10 | 0.03 |
| JP | 0.13 | 0.29 | -0.03 | **-0.37** | 0.18 |

**Table 8.** Pearson correlations between proposed instrument and BFI

## 3 Discussion

One of the identified problems in the software engineering community is the improper use of personality tests or wrongly selected assessments where many tests in the field are not conducted appropriately [22]. To cope with such issues, we created a questionnaire with situation-oriented questions based on the content of software engineering context. The goal of the assessment device (i.e. questionnaire with 70-items), which was designed for software practitioners, was to measure their personality characteristics on a newly formed MBTI compatible scale. Consequently, the developed questionnaire was tested over different groups of practitioners including expert reviewers, a pilot group of university students, a large sample of software practitioners, etc.

There were several reasons that the proposed instrument was found to have acceptable levels of the content validity, reliability and construct validity. First, internal consistency reliability coefficient (as measured by Cronbach's alpha)

found satisfactory for the entire questionnaire with different sample sets. Secondly, factor analysis indicated that the four-factor model has high item-scale correlations, which was found to be strong evidence for construct validity. Thirdly, to improve the content validity, problematic questions were analyzed, identified, and revised by expert reviews. After having a number of iterations, the high values obtained for CVIs were considered for evidence of content validity. Lastly, the correlative results from 15 participants for both the proposed scale and BFI was investigated. Hence, further studies should aim to compare two scales. Ultimately, future research with more participants should therefore concentrate on the investigation of the assessment scale and other personality inventories.

### 3.1   Threats to Validity

This section details some potential threats to validity and the methods we used to address them. To deal with internal validity problems, during the time between the two pilot studies, there was no outside event that might affect participants'; therefore, we confirmed that there were no observable change that may potentially affect the results. Secondly, participants were exposed to the same questions, therefore we did not observe a testing that might potentially affect or threaten the internal validity. In other words, during the experiments, we did not change our survey instrument (i.e. measuring device), which could potentiality be a threat to validity.

In the sixth step of the work, for different personality assessments, we built *within-participants design* in which we used the participants from the same group to take measures for the two attempts. One advantage of this work is that when the same participants contribute to the same conditions, it increases the chance of having statistical significance [23].

Finally, the construct validity shows the ability of an instrument to measure the operational form of a construct (e.g. extroversion, introversion), which it was built to measure. To investigate this phenomenon, initially construct validity of the questionnaire was comprehensively discussed by using experts from several fields (see Table 3). Secondly, we corroborated with a selected group of participants to review their personality traits (i.e. how accurate their traits were described), and latter several discussions were conducted with experts to systematically investigate the validity of the measurement scale. Notwithstanding the great care that has been taken to raise the reliability and utility of our work to date, future valuable research should focus on further evaluations of the effectiveness of the instrument.

## 4   Conclusions and Future Work

This study set out to determine the personality profiles of software practitioners by using a context dependent questionnaire. The questionnaire is based on the concept of situational context cards [24], which relies on the notion of personality that is a product of personal choices, thoughts, and opinions. Although

personality measurements could show a spectrum, which are not rigid and may consist a wider margin of variance, personality profiling is still a useful technique to understand practitioners strengths and weaknesses against particular situations [25]. However, traditional MBTI-based psychometric tests are not context oriented [11]. To deal with this issue, in this research, we formulate a context dependent approach to reveal the personality type of individuals.

This paper has shown that it is now possible to construct a context specific MBTI assessment. It potentially improves the participants engagement encouraging individuals to reveal their personality types.. The empirical findings in this study provide a new understanding of the personality profiling process, which is more context specific, tangible, and therefore obtain more concrete results from the participants. Most importantly, however, our approach explores the fabric between the actions of individuals and social landscapes of software development teams. It should therefore be possible to establish some structural improvements for a software team, based on the fact that the quality of organizational production relies on the structure of the organization [26].

The vision of this novel approach and its implementation in software development organizations can provide a way to explore the effects of personality types on team compositions where this information can be used to investigate effective team configurations. Our next goal is to dynamically portray the personality traits of an individual for designing an optimal team structure using an extended implementation of the assessment model. Such an approach should be designed to illustrate the social structure of software teams as a whole based on job roles [27], and its members' personality types.

## Acknowledgments

## References

1. Stellman, A., Greene, J.: Applied software project management. O'Reilly Media (2005)
2. Acuna, S.T., Juristo, N., Moreno, A.M., Mon, A.: A Software Process Model Handbook for Incorporating People's Capabilities. Springer-Verlag (2005)
3. Schwalbe, K.: Information Technology: Project Management. Cengage Learning (2010)
4. Lui, K.M., Chan, K.C.C.: Software Development Rhythms. John Wiley & Sons (2008)
5. Mazni, O., Syed-Abdullah, S., Hussin, N.: Analyzing personality types to predict team performance. In: Science and Social Research (CSSR), 2010 International Conference on, IEEE (2010) 624–628
6. Lewis, T., Smith, W.: Building software engineering teams that work: The impact of dominance on group conflict and performance outcomes. In: Frontiers in Education Conference, 2008. FIE 2008. 38th Annual, IEEE (2008) S3H–1

7. Su-li, Z., Ke-fan, X.: Research on entrepreneurial team members' personality traits influence on group risk decision-making. In: Management Science and Engineering (ICMSE), 2010 International Conference on, IEEE (2010) 937–942
8. Capretz, L., Ahmed, F.: Making sense of software development and personality types. IT Professional **12** (2010) 6–13
9. Kerth, N., Coplien, J., Weinberg, J.: Call for the rational use of personality indicators. Computer **31** (1998) 146–147
10. Kline, P.: The handbook of psychological testing. Psychology Press (2000)
11. Kaluzniacky, E.: Managing psychological factors in information systems work: An orientation to emotional intelligence. Information Science Publishing (2004)
12. John, O.P., Donahue, E.M., Kentle, R.L.: The big five inventoryversions 4a and 54. Berkeley: University of California, Berkeley, Institute of Personality and Social Research (1991)
13. Keirsey, D., Bates, M.: Please understand me: Character & temperament types. Prometheus Nemesis Michigan (1984)
14. Lenz, E.R.: Measurement in nursing and health research. Springer Publishing Company (2010)
15. Joreskog, K., Sorbom, D.: LISREL 8: users reference guide. Lincolnwood, IL: Scientific Software International. Inc (2001)
16. Biemer, P., Lyberg, L., Wiley, J.: Introduction to survey quality. Wiley Series in Survey Methodology. Wiley (2003)
17. Presser, S.: Methods for Testing and Evaluating Survey Questionnaires. Wiley Series in Survey Methodology. Wiley-Interscience (2004)
18. Hansen, M., Hurwitz, W., Pritzker, L., of the Census, U.S.B.: The estimation and interpretation of gross differences and the simple response variance. Bureau of the Census (1963)
19. Cohen, J.: A coefficient of agreement for nominal scales. Educational and psychological measurement **20** (1960) 37–46
20. Varona, D., Capretz, L., Piñero, Y.: Personality types of cuban software developers. Global Journal of Engineering Education **13** (2011)
21. McCrae, R.R., Costa, P.T.: Reinterpreting the myers-briggs type indicator from the perspective of the five-factor model of personality. Journal of personality **57** (1989) 17–40
22. McDonald, S., Edwards, H.: Who should test whom? Communications of the ACM **50** (2007) 66–71
23. Jackson, S.: Research methods and statistics: A critical thinking approach. Wadsworth Publishing Company (2011)
24. Yilmaz, M., O'Connor, R.V.: A software process engineering approach to improving software team productivity using socioeconomic mechanism design. ACM SIGSOFT Software Engineering Notes **36** (2011) 1–5
25. Yilmaz, M., O'Connor, R.: Towards the understanding and classification of the personality traits of software development practitioners: Situational context cards approach. In: Software Engineering and Advanced Applications (SEAA), 2012 38th EUROMICRO Conference on, IEEE (2012) 400–405
26. Yilmaz, M., OConnor, R., Collins, J.: Improving software development process through economic mechanism design. In: Proceedings of the 17th European Systems and Software Process Improvement and Innovation (EuroSPI 2010). Volume 99., Springer-Verlag (2010) 177–188
27. Yilmaz, M., OConnor, R., Clarke, P.: A systematic approach to the comparison of roles in the software development processes. Software Process Improvement and Capability Determination (2012) 198–209