

Investigating the Impact of Team Formation by Introversion/Extraversion in Software Projects

V. Garousi, and A. Tarhan

Abstract—Human factors have an important effect on performance of software teams and resulting software products. One of the seldom-studied aspects of human factors is the effect of personality-based team formation on team cohesion and quality of the software product. In this study, we investigate the above effect by conducting an exploratory case study during a term-long undergraduate software engineering course containing a project component with 50 undergraduate students. We grouped the students based on the social-interaction dimension (introversion/extraversion) of the well-known Myers–Briggs Type Indicator (MBTI) personality assessment model. We then collected the relevant metrics to explore/analyze the two parameters of interest in our study: team cohesion, and project grade as an indicator of project output (i.e. resulting product quality). Our results show that there is some (although weak) relationship between the team formation scheme (based on either introversion or extraversion) with group performance and project grade. The results also show that mixed grouping of personality types has no significant effect on team cohesion but is advantageous in achieving higher project grades especially for people with low GPAs.

Index Terms—human factors, team formation, MBTI, team cohesion, product quality, software projects, empirical study.

I. INTRODUCTION

Software Engineering (SE) is a team activity by nature, and human and social factors have a strong impact on the success of any SE endeavor and the software product developed by software teams [1]. In the pursuit of more effective and efficient software development, software teams must be composed of people who work well together. How to properly form these teams, the interaction between team members, and how individual personalities influence performance and software quality, have been among the important concerns in the SE field from the 1960s to the present day [2]. Many leading figures in the field have claimed that it is fundamentally people that make the difference between success or failure of software projects [3].

V. GAROUSI is with the Information Technology Group of Wageningen University, the Netherlands (e-mail: vahid.garousi@wur.nl).

A. TARHAN is with Department of Computer Engineering, Hacettepe University, Ankara, Turkey (e-mail: atarhan@hacettepe.edu.tr).

Manuscript received August 25, 2017; accepted Nov 16, 2017.

DOI: [10.17694/bajece.419645](https://doi.org/10.17694/bajece.419645)

Building effective software teams that would lead to project success, however, is not trivial [2]. Understanding human aspects in SE teams is crucial because having the right people in a team can *make* or *break* a project. Thus, there is a need to explore factors that bind team members and understand the elements that enable effective team performance. In this context, various factors such as personality types and skill levels should be taken into account. While there exist a large body of research in this area, e.g. [4-12], there is a need for more empirical evidence and in-depth studies which look into each personality-related factors in more detail, e.g., introversion and extraversion.

In this study, we aim at assessing the impacts of personality-based team formation on team cohesion and project output, from the viewpoints of researchers and practitioners. We conducted an exploratory case study during a term-long undergraduate SE course containing a project component with 50 students. We first assessed personality types using the widely-used Myers–Briggs Type Indicator (MBTI) [13, 14], which is the most commonly used model in SE literature [15]. For team formations, we considered the introversion/extraversion dimension of MBTI. We then investigated the effects of team formation on team cohesion and project output. The results of our case study provide insights for practitioners and can be useful when building software teams.

The remainder of this paper is organized as follows. Section II discusses the background and related work. Section III describes our research method. Section IV presents the results of the study. Section V summarizes the findings, implications and limitations of our study. Finally, Section VI concludes this study and states the future work directions.

II. BACKGROUND

A. Team Related Factors

The most related body of work to our study are the empirical studies about team-related factors in SE. From the large set of such studies, we have sampled a list as shown in Table I. For each study, we show the publication year, paper title, and the independent and dependent variable(s) studied in the study.

From the list of possible independent and dependent variables that are worthy of investigation, some have been studied in the previous work as listed in Table I. In this study, we focus on personality-based team formation as the independent variable and team cohesion and project output as dependent variables. To the best of our knowledge, our study is the first one focusing on this particular combination of independent and dependent variables.

TABLE I. EMPIRICAL STUDIES IN SE STUDYING TEAM-RELATED FACTORS

Ref. & Year	Paper title	Independent variable(s)	Dependent variable(s)
[4] 2005	Examining team cohesion as an effect of software engineering methodology	Software engineering methodology	Team cohesion
[5] 2006	A follow up study of the effect of personality on the performance of software engineering teams	Personality	Team performance
[6] 2009	How do personality, team processes and task characteristics relate to job satisfaction and software quality?	Personality, team processes and task characteristics	Job satisfaction and software quality
[7] 2010	Analyzing personality types to predict team performance	Personality types	Team performance
[9] 2010	Software engineering group work: personality, patterns and performance	Personality of group members	Using design patterns, learning achievements
[8] 2013	A worked example of the relations between personality and software team processes	Personality	Team process
[10] 2014	A mixed methods investigation of ethnic diversity and productivity in software development teams	Ethnic diversity	Productivity, innovation and problem solving
[11] 2014	A replicated quasi-experimental study on the influence of personality and team climate in software development	Team cohesion and conflict	Team performance
[12] 2015	Are team personality and climate related to satisfaction and software quality? Aggregating results from a twice replicated experiment	Team personality and climate	Satisfaction and software quality

TABLE II. DIMENSIONS IN MBTI PERSONALITY ASSESSMENT

Introversion/ Extraversion (I/E)	<i>The extraverted types</i> learn best by talking and interacting with others; and by interacting with the physical world, they can process and make sense of new information. <i>The introverted types</i> prefer quiet reflection and privacy; and information processing occurs as they explore ideas and concepts internally.
Sensing/ Intuition (S/N)	<i>Sensing types</i> enjoy a learning environment in which the material is presented in a detailed and sequential manner. They attend to what is occurring in the present, and can move to the abstract after they have had the experience. <i>Intuitive types</i> prefer a learning atmosphere in which an emphasis is placed on meaning and associations; they value insight higher than careful observation, and naturally recognizes patterns in work.
Thinking/ Feeling (T/F)	<i>Thinking types</i> desire objective truth and logical principles and are natural at deductive reasoning. <i>Feeling types</i> place an emphasis on issues and causes that can be personalized while they consider other people's motives.
Judging/ Perceiving (J/P)	<i>Judging types</i> thrive when information is organized and structured, and they are motivated to complete assignments in order to gain closure. <i>Perceiving types</i> flourish in a flexible learning environment in which they are stimulated by new and exciting ideas.

B. MBTI Personality Assessment

The Myers–Briggs Type Indicator (MBTI) is a popular tool for personality assessment, which was developed based on the theories of Carl Jung [16]. It serves as an introspective self-report questionnaire designed to indicate psychological preferences in how people perceive the world and make decisions [13, 14]. There are four dimensions in this indicator as shown in Table II. MBTI has been widely used in different research communities, e.g., social sciences, psychology and SE [15]. Various studies have appeared on the usage of MBTI and other personality tests in SE, e.g., [3, 15, 17]. According to a systematic literature review (SLR) on personality assessment in SE [15], MBTI is the most commonly used model in the SE literature. For all the above reasons, in this study, we selected MBTI as the personality assessment model.

C. Team Building and Personality Types

There are many discussions and studies in other domains which report collaborations among extraverts and introverts could be challenging [18, 19], e.g.: “*Extraverts can think that introverts are slow, have few ideas to share and are unemotional. They interpret those calm faces as meaning introverts lack in emotion and passion. Introverts think that extraverts are shallow because they talk a lot. Not being direct and concise can be seen as lacking depth.*” [19].

It was argued in [2] that personality-type analysis could help take the guesswork out of putting together a high-performance software project team. The authors invited 92 Information Systems (IS) professionals from 20 software development teams in Hong Kong to complete a questionnaire-based survey. The surveys showed how team leaders scored on the information gathering dimension (sensing/intuitive) had a significant impact on team performance. Only the decision-making dimension (thinking/feeling) of the systems analyst personality had a significant influence on team performance. Only the social-interaction dimension (introversion/extraversion) of the programmer personality was strongly related to team performance. Among the conclusions were that it was unnecessary to have diversity of personalities among team members (excluding team leader) due to the fact that members needed to perform multiple tasks of the software development life cycle and that heterogeneity was not good for all phases.

The study in [20] examined the relationships between the ‘Big Five’ personality factors (conscientiousness, extraversion, neuroticism, agreeableness, and openness to experience) and objective team performance, and derived implications for selecting successful product teams. Successful teams were characterized by higher levels of general cognitive ability, higher extraversion, higher agreeableness, and lower neuroticism than their unsuccessful counterparts.

The study in [21] proposed a formal model for assigning human resources to teams in software projects. Using the Delphi method, the authors proposed a set of software project roles and competencies. Psychological tests and data mining tools identified useful rules for forming software project teams. These were used to build a formal model, which was later built into a tool that automatically calculated role assignments. This decision-support tool was claimed to help managers in assigning people to roles and forming software teams. The model was validated by assignment scenarios in two software development organizations.

The study in [22] presented a mix-method replicated study for team building in software projects. The findings indicated that carefully selecting team members for software teams was likely to positively influence the projects in which these teams participate. Besides, it seemed that the type of development method could moderate (increase or decrease) this influence.

The study in [23] discussed a comparison of the performance of student groups formed randomly, with those formed by using the learning styles questionnaire. The study found no significant differences in the performances of these two sets of groups, for which it discussed several possible reasons.

D. Team Cohesion

One of the most studied team variables in literature is “team cohesion”. Team cohesiveness is the degree to which team members like each other, identify themselves positively with the team and want to remain with its members [24]. It reflects the degree of attraction among group members. A study of cohesiveness is considered essential for understanding group dynamics in teams. Two meta-analyses in the psychology discipline [25, 26] have reported a positive relationship between cohesiveness and performance. According to these studies, cohesive teams demonstrate increased collective efficacy and greater team success. Furthermore, cohesive team members are less anxious, more satisfied, have higher self-esteem, conform to group norms, make personal sacrifices for the team, share responsibility for team failure and are less likely to indulge in social loafing.

III. RESEARCH METHOD

A. Goal and Research Questions (RQs)

The goal of our study was to conduct an exploratory evaluation on impacts of personality-based team formation on team cohesion and also quality of project output (software). To focus the study on one independent variable and prevent the impact of more than one independent variables (the so called “confounding bias”), we focused on only one dimension of the MBTI – the social-interaction dimension (introversion and extraversion). Based on the stated goal, we raised the following two research questions (RQ):

- RQ 1 - What are the impacts of personality-based team formation on team cohesion?
- RQ 2 - What are the impacts of personality-based team formation on project output (i.e. resulting product quality)?

B. Research Design

We designed our research approach by adapting the Goal, Question, Metric (GQM) methodology [27]. We replaced the questions with RQs, and identified independent and dependent variables as the metrics to be used in our research to answer the RQs. The design that we developed is shown Table III.

TABLE III. GQM DESIGN OF OUR RESEARCH APPROACH

Goal: To conduct an ‘exploratory’ evaluation on impacts of personality-based team formation on team cohesion and project output		
RQ1: What are the impacts of personality-based team formation on team cohesion?		
Independent var.	M1: MBTI social interaction personality type (I/E)	
	M2: Students’ grade point average (GPA)	
Dependent var.	M3: Team cohesion morale index (TCMI)	
RQ2: What are the impacts of personality-based team formation on project output?		
Independent var.	M1: MBTI social interaction personality type (I/E)	
	M2: Students’ grade point average (GPA)	
Dependent var.	M4: Project grade (as an indicator of project output)	

For personality-based team formation, we needed a suitable metric for assessing students’ personality types. We instructed the students in the beginning of the semester to take MBTI using a free online test [14] for identifying their own personality types of social interaction dimension (introversion/extraversion). We also gathered students’ latest cumulative

grade point average (GPA) in their program, and treated the GPA as indicators of their technical abilities. We used MBTI social interaction personality types and GPAs of students as independent variables in our research.

Metrics for the dependent variables included the ones for measuring the team cohesion and the project output. For quantitatively measuring team cohesion, we searched in both the formal and the grey literature, and selected a rubric set [28] developed by an Agile practitioner and coach. This rubric is used by Agile practitioners to quantitatively measure team morale in Agile teams in the industry. The rubric, which is shown in Table IV, has been developed using the rigorous foundations from the psychology literature [29], and consists of eight questions. The answer of each question is based on a 5-point Likert scale as follows: {1: Very low, 2: Low, 3: Average, 4: High, 5: Very high}. To quantitatively calculate the team morale of an individual member, the average value of the scores on the eight questions is calculated and set as the Team Cohesion and Morale Index (TCMI). An example calculation is shown in Table IV.

Note that while Likert scale is originally an ordinal scale, analyzing Likert scales as interval values (and calculating average based on such data) is possible when the sets of Likert items can be combined to form indexes, with the caveat (assumption) that “*this combination forms an underlying characteristic or variable*” [30]. Also, we ensured precise wordings for the five response levels above to clearly imply “*a symmetry of response levels about a middle category*” [31]. Therefore, equal spacing of response levels was clearly indicated, and the argument for treating it as interval-scale data was supported [31].

TABLE IV. 8-QUESTION RUBRIC USED TO MEASURE TEAM COHESION AND MORALE (ADOPTED FROM [28])

#	Question	Sample values
1	I am enthusiastic about the work that I do for my team	1: Very low
2	I find the work I do for my team of meaning and purpose	4: High
3	I am proud of the work that I do for my team	3: Average
4	To me, the work that I do for my team is challenging	2: Low
5	In my team, I feel bursting with energy	5: Very high
6	In my team, I feel fit and strong	4: High
7	In my team, I quickly recover from setbacks	3: Average
8	In my team, I can keep going for a long time	1: Very low
	Team Cohesion and Morale Index (TCMI)=	Avg.=2.8

For quantitatively measuring project output, we used project grade as an indicator of resulting project output. We calculated project grades based on the grades of deliverables planned and submitted by students throughout project life cycle. Students delivered project artifacts at five milestones: (0) Vision and project plan, (1) Requirements document, (2) Design document, (3) Demo of the prototype version of the software, and (4) Final software product. Students were asked to submit team cohesion (morale) via an online questionnaire in each milestone (0-4). Project teams had to explain their team work and the work of each student individually in project reports which were then used for marking the works of teams and the students. The teaching team assessed the students’ works in each delivery, and used the following rubric to reduce subjectivity in marking: Functional quality (% of test cases passed), code readability,

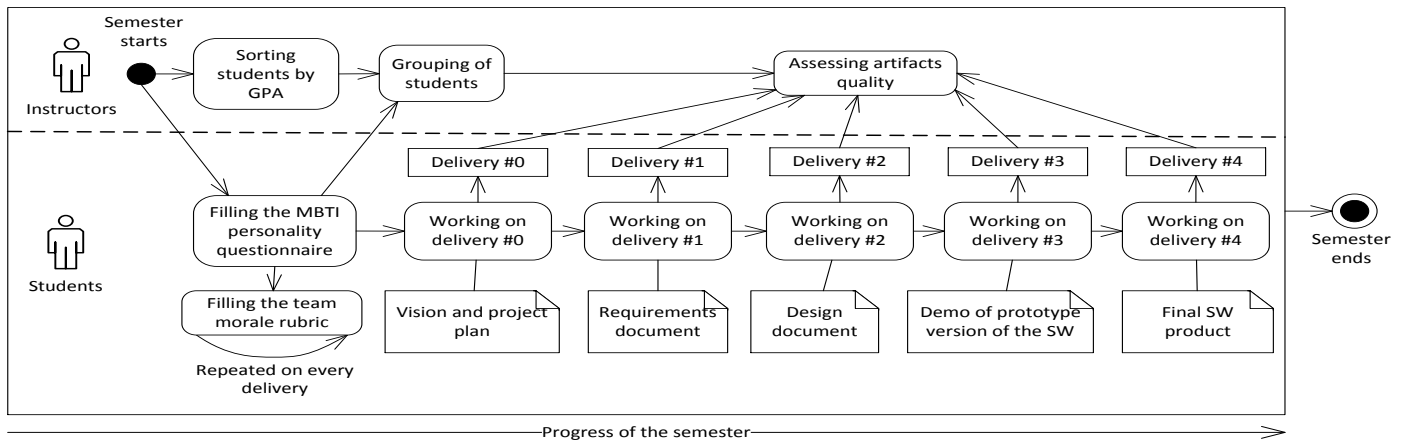


Fig. 1. Activity diagram showing the planning and execution of the empirical study.

and extent and quality of documentation. Although the authors had intended to use a more detailed rubric for evaluating deliveries, they had to simplify the rubric to include several key factors above due to shortage of human resources.

Fig. 1 depicts an activity diagram showing the planning and execution stages of our empirical study. For its design and execution, we took into consideration the recommendations on using students in empirical studies (e.g., [32, 33]) and also received ethics approval from Hacettepe University.

The teaching team consisted of two instructors (the co-authors of this paper) plus two teaching assistants (TAs). As shown in Fig. 1, in the beginning of the semester, the instructors sorted the students by their GPAs and used the MBTI personality data to group the students. The approach that we took for grouping is explained in the next subsection.

C. Study Subjects and Team Formation Approach

The study was conducted in the context of ‘Software Engineering Laboratory’ course, which is the practical counterpart of the 3rd year ‘Software Engineering’ course, in Hacettepe University’s Department of Computer Engineering. During the Spring 2016 offering of the course, in which the study was conducted, the course had exactly 50 students.

An important issue was to decide how to group students (i.e. form teams). As per the study’s goal (impacts of personality-based team formation by social-interaction dimension of the MBTI), we sorted the students by their GPAs and then grouped each three students into one team such that, in a controlled manner, a group with the closest GPAs would have all introvert members, another group would have all extravert students, and another one having a mix of introverts and extraverts. Grouping would continue from students with higher GPAs towards those with lower GPAs until every student belonged to a group. This grouping mechanism would yield a set of groups with similar GPAs in which the only differentiating factor would be the extraversion and introversion attitudes.

As shown Fig. 1, we instructed all the students in the beginning of the term to take an online MBTI test [14] and send their results to the instructors. Once we had the MBTI assessment results (types) and the GPAs, we used the grouping approach discussed above to form the groups. Fig. 2 shows the

results of grouping process. We set the group sizes to three students, except the very last group (which had five students). As a result, 16 groups were formed, shown as #1 ... #16 in the figure. As per our grouping approach, all three members of group #1 were extraverts (labeled as ‘All Ext’ in Fig. 2). All three members of group 2 were introverts. Group 3 was a mix of extraverts and introverts, etc.

Based on the MBTI data, at the end, we had three all-introvert groups. Nine groups had all extraverts, and four groups were mixes of introverts and extraverts. We would have liked to have a balanced mix of the three group types, but the MBTI data of students did not allow this (i.e., we did not have as many introverts as we wanted to put in the groups).

Fig. 3 shows a dot-plot of the distribution of the GPA values of the 50 students in the class. GPA values were out of 4 and corresponded to the performance of the students in their previous school terms. The values were taken from the student records. The mean (average) was 2.55. The minimum and maximum were 1.11 and 3.45, respectively.

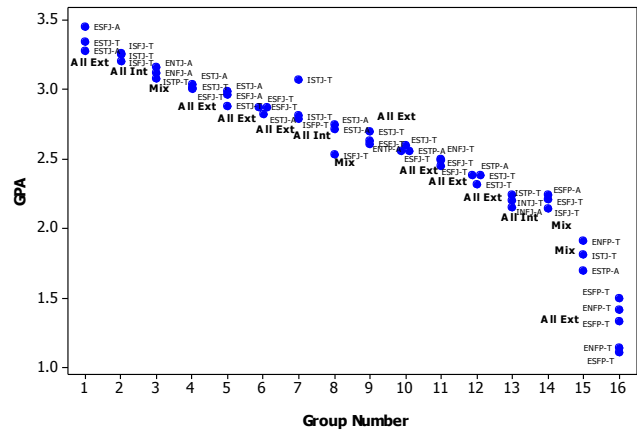


Fig. 2. GPAs and MBTI social interaction types of the students in 16 groups

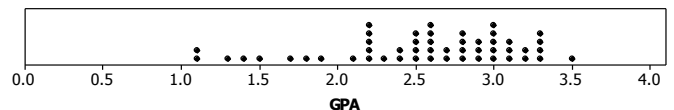


Fig. 3. Dot-plot of the GPA values of the 50 students in the class

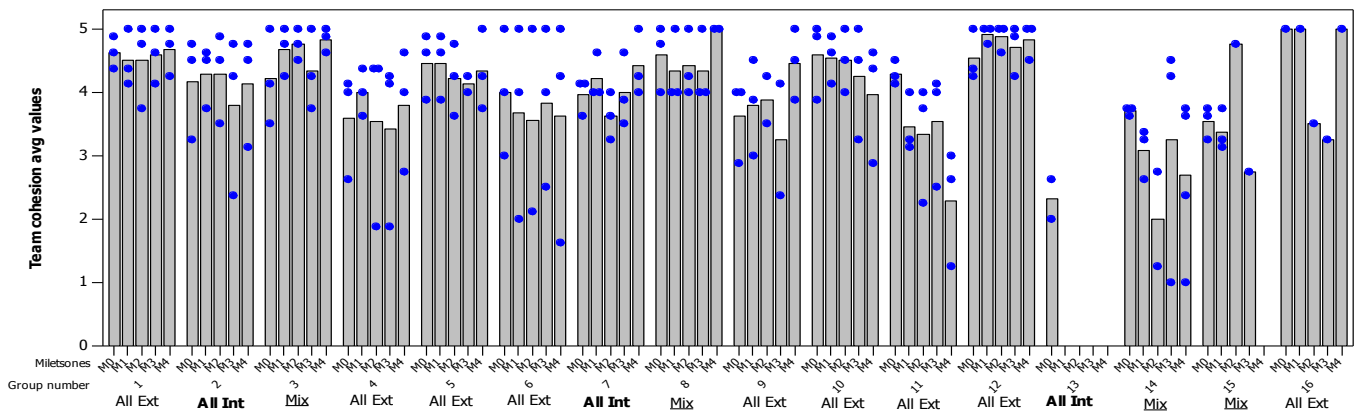


Fig. 4. Team cohesion average values reported by each group member (student) in each milestone

D. Study Objects and Project Development Context

The project was to develop an online library management software. Students were provided with the high-level requirements of the system written in English, and in a UML (Unified Modeling Language) use-case diagram.

Students were asked to use the Open Unified Process (OpenUP) development process and its artifacts' templates [34]. As shown in the design of the empirical study (Fig. 1), the development project had five milestones and students submitted various software artifacts (documentation or code) in each of the steps, as per the OpenUP's specifications. We asked for artifacts for the following phases: requirements, architecture, implementation, testing, and project management.

For requirements and design stages, OpenUP requires modeling by using UML. To establish consistency in the entire class among all student groups, students were asked to use the Visual Paradigm UML tool [35].

IV. RESULTS

A. RQ1: Team Formation and Team Cohesion Morale Index

Our rationale behind RQ1 was to assess the implications and outcomes of team dynamics and to assess the impacts of team formation (if any) on team cohesion and morale as measured by Team Cohesion and Morale Index (TCMI) metric. Fig. 4 shows, as an individual-value plot, the team cohesion average values reported by each group member (student) in each milestone. The bars show the average values of the individual values for each milestone, e.g., M0 thru M4. Data for group 13 was not available since the group decomposed soon after the term had started. Please note that the team who abandoned the class was one of the teams with the lowest GPAs, which explains their decision to drop the course. Only two TCMI values for M0 for this group were reported. This did not lead to a negative effect on our case study since our group formation took into account academic success, and our design had a preventive nature against such occurrences as previously mentioned. Recall from Section III.C that groups were sorted by descending order of GPAs, e.g., members of group #1 had the highest GPAs and those in group #16 had the lowest GPAs.

For ease of review and analysis, we have also included the types of groups (either all extraverts, all introverts, or mixed) in Fig. 4 (below the group numbers). As we could observe, grouping based on introversion/extraversion 'alone' did not

have any noticeable impact on team cohesion, as groups with all extraverts, all introverts, or mixed all reported different levels of the TCMI measured, regardless of group formation types. One expectation in this context could have been that, in groups with homogeneous (compatible) team members (all extraverts or all introverts), TCMI measure would be higher than in groups with mixes of extraverts and introverts, since mixed groups could have higher chances for arguments and disagreements, thus leading to lower TCMI measure.

We found no significant correlation between the social-interaction dimension (introversion/extraversion) and team cohesion (i.e. TCMI values). This observation is similar to the findings of the study in [11] in which no significant correlation between the extraversion personality factor and team satisfaction was found.

We also investigated whether there was any correlation between TCMI values reported by each student and her/his GPA (i.e., whether students with higher GPAs felt better team cohesions). Fig. 5 shows the scatterplot of these values for all students. The Pearson correlation of the two datasets is 0.24 (p-Value = 0.12) – thus, showing a weak correlation, meaning that for a student with higher technical capabilities, it would be expected for her/him to have a higher perception of team morale and team cohesion feelings; and vice versa.

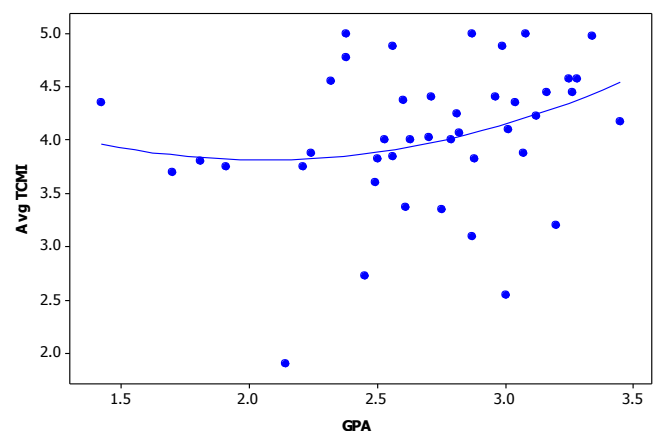


Fig. 5. Scatterplot of TCMI values reported by each student and her/his GPA

B. RQ2: Team Formation and Project Grade

As the response to RQ2, we discuss the impacts of team formation on project grade as an indicator of resulting project

output. Fig. 6 shows the individual-value plots of (a) GPA and (b) project grades with respect to team formation, i.e., all extraverts (All Ext) – 29 students, all introverts (All Int) – 9 students, and mixed (Mix Ext-Int) – 12 students.

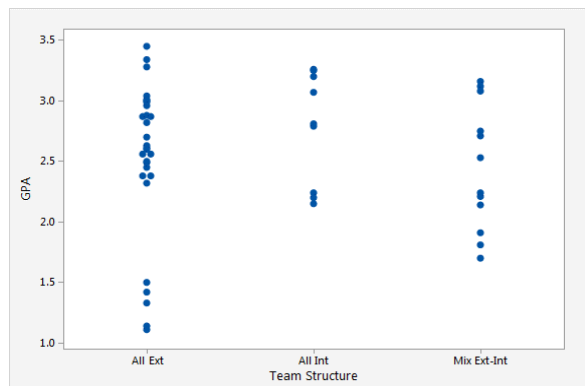
As Fig. 6 (a) shows, students in the teams of all-extraverts mostly had high GPAs (between 2.3 and 3.45), though a small number of such students had GPAs below 1.5. It is also seen from the figure that students in the teams of all-introverts had GPAs above 2.15. When it comes to students in the mixed teams of extraverts and introverts, we see that GPAs were distributed between the values of 1.7 and 3.16. In addition, the students in all-extraverts-teams had values on the edges of the GPA scale while the range for the students in all-introverts-teams was smaller in variance (between 2.1 and 3.2).

Fig. 6 (b), on the other hand, represents the individual-value plots of project grade for the three team types. For the teams having all extraverts, most of the students (75%) received grades above 65% and that four out of 29 students failed. The teams having all introverts either were very successful (having grades above 93%) or performed very poorly (in three out of nine groups). The mixed teams, interestingly, were generally

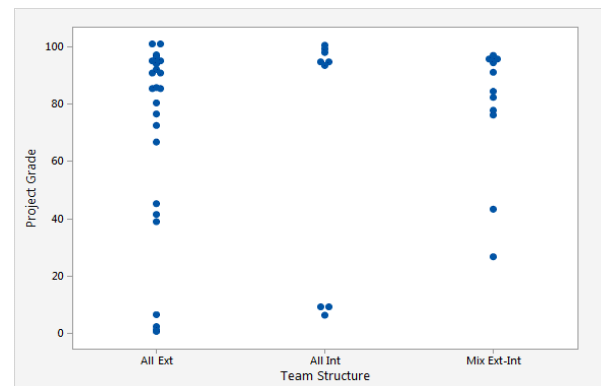
high-performers (with two exceptions) with their grades between 76% and 97%.

To evaluate the data in Fig. 6 (a) and (b) together, and to better understand the relationships between GPAs and project grades for different team types, we sketch in Fig. 7 the average values of project grades versus the average values of GPAs of the teams with respect to the three team types. The figure shows the relation between project grades and GPAs of group members per team formation type. It can be observed that the groups with lower GPAs failed except the mixed ones, i.e., having both extraverts and introverts. It seems mixed grouping of personality types worked better than discrete grouping of all extraverts or all introverts, in terms of achieving higher project grades, especially for students with low GPAs.

We also investigated the relationship between project grades and GPAs of the students by their personality type. Fig. 8 shows the distributions of data points, with their best-fit-curves, for the types of extraverts (Ext) and introverts (Int). Coincidentally, best-fit-curves are fully overlapping, denoting that there was no statistically-significant difference in project grades between the students from the two personality types (extraverts and introverts).



(a) Plot of GPA values



(b) Plot of project grades

Fig. 6. Individual-value plots of GPAs and project grades of students by team structure

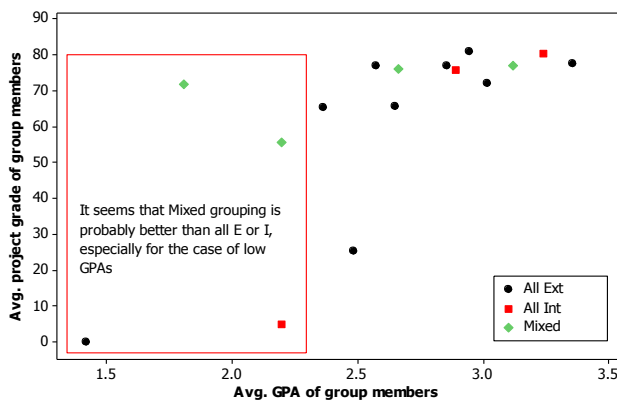


Fig. 7. Avg. project grade versus avg. GPA of teams by team formation

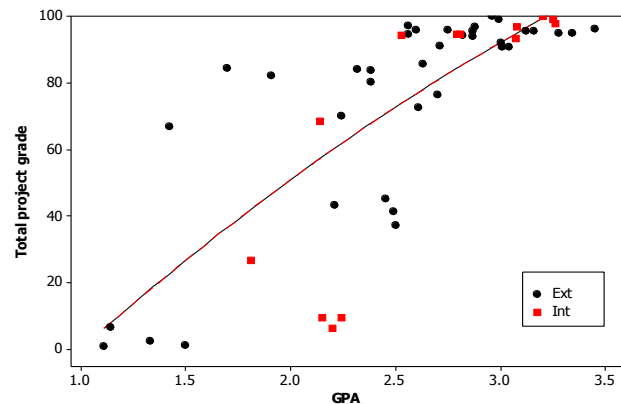


Fig. 8. Project grade versus GPA, grouped by personality types: E vs. I

V. DISCUSSION

A. Summary of Findings

RQ1 was intended to investigate the implications and outcomes of team formation on team cohesion and morale as measured by the Team Cohesion and Morale Index (TCMI) metric. The results for RQ1 showed that grouping based on the social-interaction dimension ‘alone’ (introversion vs. extraversion) did not have any noticeable impact on team cohesion, as groups with all extraverts, all introverts, or mixed types reported different levels of TCMI values, regardless of the group formation types. This observation was similar to the findings of the study [11] in that no significant correlation between the extraversion personality type and team satisfaction was found. We also examined if there was any correlation between TCMI values reported by each student and her/his GPA (i.e. if the students with higher GPAs felt better about team cohesions). We noticed a weak correlation between these two variables, possibly meaning that the higher the technical capabilities of a developer, the stronger his/her feelings of team morale and team cohesion would be; and vice versa.

RQ2 was aimed to understand the impacts of team formation on project grades as an indicator of resulting project output. The results for RQ2 showed that, for the teams having all extraverts, most of the students received grades above 65% and that only few students in such groups failed. Students in the all-introvert groups were either very successful (with grades above 93%) or performed very poorly. The mixed teams, interestingly, were generally high-performers. That is, mixed grouping of personality types worked better than discrete grouping of all extraverts or all introverts, in terms of project grades, especially for people with low GPAs (i.e. low technical abilities).

B. Potential Threats to Validity

We discuss the limitations and potential threats (construct, internal, conclusion, external) [36] to the validity of our study and the steps that we took to minimize or mitigate them in the following paragraphs.

Construct validity is concerned with issues that to what extent the study truly represented the theory behind it [36]. The potential issues in this regard were whether we properly conducted personality-based team formation, and actually measured team cohesion and project grades. Adapting the GQM methodology [27] for our research design and standardizing the metrics and the instruments used in this study, we addressed those issues and minimized the associated threats. However, threats might have remained regarding the variability of MTBI test results depending on the mood of the students while answering the questions, and the percent rating scheme of the questionnaire, e.g., introversion/extraversion scores could be close (e.g. 49% vs. 51%) or far apart (e.g. 1% vs. 99%). Also, the team formation approach that we used based on the students’ GPAs might be considered as another threat, as we put the best students together, average students together, and not-so-good students together. Still, we adopted this grouping mechanism because it resulted in a set of groups with similar GPAs, which was important to keep the social interaction dimension (i.e. extraversion/introversion) the only differentiating factor in our research design.

Internal validity reflects the extent to which a causal conclusion based on a study is warranted [36]. To prevent confounding bias, we focused only on the social-interaction dimension (introversion/extraversion) of the MBTI model, and thus prevented the likely impact of more than one independent variables. In terms of selection bias, the subjects of the study were composed of 50 undergraduate students who had enrolled in the ‘Software Engineering Laboratory’ course. To prevent any negative influence, we considered the recommendations on using students in empirical studies (e.g., [32, 33]) in the design of the study, and had ethics approval from our university. While the subjects (i.e. the students) were not yet software engineers, they had very similar profiles. To reduce possible variability in team activities and project deliverables, the students followed the basic life-cycle that we tailored from OpenUP and used its artifacts’ templates [34]; and also used a popular UML modeling tool [35].

Conclusion validity of a study deals with whether correct conclusions are reached through rigorous and repeatable treatments [36]. To reduce the bias in reaching conclusions for each research question, we relied on statistical analysis. Thus, interpretation of the findings and implications of our research depends on statistical significance and are strictly traceable to data. In addition, by careful definition of evaluation process, its outputs and their grading rubrics, we enabled valid and repeatable investigation of the RQs.

External validity is concerned with to what extent the results of this study can be generalized [36]. The study was done in a single university course with only 50 undergraduate students formed into 16 groups. It provides a limited voice of evidence from a small data set, and therefore generalizing its findings is not possible. Though our study added to the body of evidence on this area, replications of it in other contexts would be needed to increase generalizability of our findings.

VI. CONCLUSIONS AND FUTURE WORK

In this study, we investigated the effect of personality-based team formation (based on social-interaction dimension of MBTI, i.e. introversion/extraversion) on team cohesion and project output (i.e. resulting product quality). We conducted an exploratory case study during a term-long software engineering course containing a project component with 50 students. We grouped the students using the Myers–Briggs Type Indicator (MBTI) personality assessment model; and collected data to explore the team cohesion as measured by team cohesion and morale index, and project grade as an indicator of project output. Our results showed that there was no relation between team formation types and team cohesion, and that some (although weak) relationship existed between the formation schemes and group performance and resulting project output.

Our study provides a limited voice of evidence from a small data set. While our study added to the body of evidence in this area, it also highlighted the very complex nature of human characteristics and its manifestation in team formation and likely results. As researchers, we need to look in further depth into team dynamics and human aspects in software teams. We could consider other dimensions of personality types (i.e., sensing/intuition, thinking/feeling, judging/perceiving) in addition to social interaction dimension (i.e. extraversion/introversion), and their influence on team cohesion and

resulting project output. The findings from such research, including our investigation, might provide insights for practitioners who want to build teams to evaluate and increase the efficiency of their software teams.

Our future work directions include: (1) investigating to see whether Agile teams have a higher team morale than other teams (e.g., working in Waterfall); (2) investigating the effects of other MBTI dimensions on team cohesion and project output; (3) investigating the effects of uniform teams (all extraverts or all introverts) on development activities and if the performance differs in various SDLC phases: Analysis, design, implementation, and testing; and, (4) investigating to see whether more homogeneous teams (in terms of personality) are more suitable for software development compared to less homogeneous teams.

ACKNOWLEDGMENT

This study received ethics approval from Hacettepe University in 2015. The authors thank to Tuğba Erdoğan and Pelin Canbay, the TAs of the course, in their assistance in delivery of the laboratory course and marking the project deliverables.

REFERENCES

- [1] T. DeMarco and T. Lister, *Peopeware: productive projects and teams*. Dorset House Publishing Co., Inc., 1987, p. 188.
- [2] N. Gorla and Y. W. Lam, "Who should work with whom? building effective software project teams," *Commun. ACM*, vol. 47, no. 6, pp. 79-82, 2004.
- [3] L. T. Hardiman, "Personality Types and Software Engineers," *Computer*, vol. 30, no. 10, pp. 10-10, 1997.
- [4] C. A. Wellington, T. Briggs, and C. D. Girard, "Examining team cohesion as an effect of software engineering methodology," *SIGSOFT Softw. Eng. Notes*, vol. 30, no. 4, pp. 1-5, 2005.
- [5] J. Kam and T. Cowling, "A follow up study of the effect of personality on the performance of software engineering teams," presented at the Proceedings of the ACM/IEEE international symposium on Empirical Software Engineering, 2006.
- [6] S. T. Acuña, M. Gómez, and N. Juristo, "How do personality, team processes and task characteristics relate to job satisfaction and software quality?," *Information and Software Technology*, vol. 51, no. 3, pp. 627-639, 2009.
- [7] O. Mazni, S.-L. Syed-Abdullah, and N. M. Hussin, "Analyzing personality types to predict team performance," in *International Conference on Science and Social Research (CSSR)*, 2010, pp. 624-628.
- [8] F. Q. B. da Silva, S. S. J. O. Cruz, T. B. Gouveia, and L. F. Capretz, "Using Meta-ethnography to Synthesize Research: A Worked Example of the Relations between Personality and Software Team Processes," in *ACM IEEE International Symposium on Empirical Software Engineering and Measurement*, 2013, pp. 153-162.
- [9] D. Bell, T. Hall, J. E. Hannay, D. Pfahl, and S. T. Acuna, "Software engineering group work: personality, patterns and performance," in *Proceedings of the Annual Conference on Computer Personnel Research*, 2010, pp. 43-47, 1796921.
- [10] J. Congalton, "A mixed methods in investigation of ethnic diversity and productivity in software development teams," PhD thesis, Information Systems, Massey University, Wellington, New Zealand, 2014.
- [11] M. Gómez and S. Acuña, "A replicated quasi-experimental study on the influence of personality and team climate in software development," (in English), *Empirical Software Engineering*, vol. 19, no. 2, pp. 343-377, 2014.
- [12] S. T. Acuña, M. N. Gómez, J. E. Hannay, N. Juristo, and D. Pfahl, "Are team personality and climate related to satisfaction and software quality? Aggregating results from a twice replicated experiment," *Information and Software Technology*, vol. 57, pp. 141-156, 2015.
- [13] Constance Lynn Counselling, "Myers-Briggs Type Indicator (MBTI)" <http://www.constancelynn.com/assessments/myers-briggs-type-indicator-mbti/>, Last accessed: Nov. 2017.
- [14] NERIS Analytics Limited, "Free online personality test for Myers-Briggs Type Indicator (MBTI)," <http://www.16personalities.com/free-personality-test>, Last accessed: Nov. 2017.
- [15] S. S. J. O. Cruz, F. Q. B. da Silva, C. V. F. Monteiro, C. F. Santos, and M. T. dos Santos, "Personality in software engineering: Preliminary findings from a systematic literature review," in *Annual Conference on Evaluation & Assessment in Software Engineering*, 2011, pp. 1-10.
- [16] D. P. McAdams, *The Person: An Integrated Introduction to Personality Psychology*. Harcourt, 2001.
- [17] S. McDonald and H. M. Edwards, "Who should test whom? Examining the use and abuse of personality tests in software engineering," *Commun. ACM*, vol. 50, no. 1, pp. 66-71, 2007.
- [18] J. B. Kahnweiler, *The Genius of Opposites: How Introverts and Extroverts Achieve Extraordinary Results Together*. Berrett-Koehler Publishers, 2015.
- [19] Forbes, "Can Introverts And Extroverts Ever Work Well Together? How Opposites Can Collaborate Brilliantly," <http://www.forbes.com/sites/kathycapriano/2015/09/07/can-introverts-and-extroverts-ever-work-well-together-how-opposites-can-collaborate-brilliantly>, 2015, Last accessed: Nov. 2017.
- [20] S. L. Kichuk and W. H. Wiesner, "The big five personality factors and team performance: implications for selecting successful product design teams," *Journal of Engineering and Technology Management*, vol. 14, no. 3-4, pp. 195-221, 1997.
- [21] M. André, M. G. Baldoquín, and S. T. Acuña, "Formal model for assigning human resources to teams in software projects," *Information and Software Technology*, vol. 53, no. 3, pp. 259-275, 2011.
- [22] F. da Silva *et al.*, "Team building criteria in software projects: A mix-method replicated study," *Information and Software Technology*, vol. 55, no. 7, pp. 1316-1340, 2013.
- [23] M. Huxham and R. Land, "Assigning Students in Group Work Projects. Can We Do Better than Random?," *Innovations in Education & Training International*, vol. 37, no. 1, pp. 17-22, 2000.
- [24] M. E. Shaw, R. Robbin, and J. R. Belsler, *Group Dynamics: The Psychology of Small Group Behavior*. McGraw-Hill, 1981.
- [25] C. R. Evans and K. L. Dion, "Group cohesion and performance a meta-analysis," *Small group research*, vol. 22, no. 2, pp. 175-186, 1991.
- [26] B. Mullen and C. Copper, "The relation between group cohesiveness and performance: An integration," *Psychological Bulletin*, vol. 115, no. 2, pp. 210-227, 1994.
- [27] V. R. Basili, "Software modeling and measurement: the Goal/Question/Metric paradigm," Technical Report, University of Maryland at College Park 1992.
- [28] C. Verwijs, "Agile Teams: Don't use happiness metrics, measure Team Morale," <http://blog.agilistic.nl/agile-teams-dont-use-happiness-metrics-measure-team-morale/>, Last accessed: Nov. 2017.
- [29] L. v. Boxmeer, C. Verwijs, R. d. Bruin, J. Duel, and M. C. Euwema, "A direct measure of Morale in the Netherlands Armed Forces Morale Survey: theoretical puzzle, empirical testing and validation," in *Presented at International Military Testing Association Symposium*, 2007.

- [30] I. E. Allen and C. A. Seaman, "Likert Scales and Data Analyses," <http://asq.org/quality-progress/2007/07/statistics/likert-scales-and-data-analyses.html>, Last accessed: Nov. 2017.
- [31] Anonymous author(s), "How to Use the Likert Scale in Statistical Analysis," <http://statisticscafe.blogspot.nl/2011/05/how-to-use-likert-scale-in-statistical.html>, Last accessed: Nov. 2017.
- [32] J. Carver, L. Jaccheri, S. Morasca, and F. Shull, "Issues in using students in empirical studies in software engineering education," in *Proceedings International Software Metrics Symposium*, 2003, pp. 239-249.
- [33] M. Svahnberg, A. Aurum, and C. Wohlin, "Using students as subjects - an empirical evaluation," presented at the Proceedings of the ACM-IEEE international symposium on Empirical software engineering and measurement, 2008.
- [34] Eclipse foundation, "Open Unified Process (OpenUP) development process," <http://epf.eclipse.org/wiki/openup/>, Last accessed: Nov. 2017.
- [35] Visual Paradigm International, "Visual Paradigm tool," <https://www.visual-paradigm.com>, Last accessed: Nov. 2017.
- [36] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén, *Experimentation in Software Engineering: An Introduction*. Kluwer Academic Publishers, 2000.

BIOGRAPHIES



VAHID GAROUSI received his B.S. and M.S. degrees in Computer Engineering from Sharif University of Technology (Iran) in 2000, and the University of Waterloo (Canada) 2003. He earned his PhD in Software Engineering in Carleton University (Canada) in 2006. He is currently an Associate Professor of Software Engineering in Wageningen University, the Netherlands. Previously, he was an Associate Professor of Software Engineering in Hacettepe University in Ankara, Turkey (2015-2017) and in the University of Calgary, Canada (2006-2014). Vahid was an IEEE Computer Society Distinguished Visitor from 2012 to 2015. His research interests in software engineering include: software testing and quality assurance, model-driven development, and software maintenance.



AYÇA TARHAN received the B.S. and M.S. degrees in computer engineering from Ege University in 1995 and from Dokuz Eylul University in 1999, and the Ph.D. degree in information systems from Informatics Institute of Middle East Technical University (METU) in 2006. She was a visiting researcher in 2013-2015 in the Department of Industrial Engineering and Innovation Sciences of Eindhoven University of Technology. Her research interests include internal and external software quality, software metrics, software development methodologies, process maturity, and business process management. Since 2007, she has been working as a Lecturer and an Assistant Professor in Computer Engineering Department of Hacettepe University in Ankara, Turkey.