DETERMINING THE RELATIVE IMPORTANCE OF PERSONALITY TRAITS IN INFLUENCING SOFTWARE QUALITY AND TEAM PRODUCTIVITY

Nosheen QAMAR

Department of Computer Science and Information Technology
University of Lahore
Lahore, Pakistan
&
Department of Computer Science
National University of Computer and Emerging Sciences
Lahore, Pakistan
e-mail: `nosheen.qamar@cs.uol.edu.pk`

Ali Afzal MALIK

Department of Computer Science
National University of Computer and Emerging Sciences
Lahore, Pakistan
e-mail: `ali.afzal@nu.edu.pk`

Abstract. Software projects are almost always team efforts and successful projects involve well-formed and well-composed teams. Past studies have revealed that personality contributes to effective team composition and, therefore, project success. Yet despite its importance, only a couple of empirical studies have quantitatively evaluated the impact of personality on software quality and team productivity. Our previous study was an effort in this direction. In that study, we proposed a metric called Team Homogeneity Index and evaluated its impact on software quality and team productivity for two phases (implementation and testing) of the software development life cycle. This study is a continuation of our previous work. In this study, we replicate our experiment on three different phases of software development life cycle (i.e. analysis and design, implementation, and testing). We also determine the weights for all five personality traits using input from the industry and propose

an improved version of Team Homogeneity Index called Weighted Team Homogeneity Index. Finally, we conduct a comparative analysis of Team Homogeneity Index and Weighted Team Homogeneity Index to determine whether weights assigned to personality traits make any difference. Our findings reveal that weights do make a difference and Weighted Team Homogeneity Index is more strongly correlated than Team Homogeneity Index for almost all of the teams, especially those composed of practitioners, in the three different phases of Software Development Life Cycle.

1 INTRODUCTION

The modern society progressively demands quality and productivity in all aspects of life. In every field, including software development, people are ready to adopt new approaches to improve product quality. Software quality refers to the degree to which all attributes of a software system appropriately fulfill its requirements [15]. Software development productivity, on the other hand, is defined as the functional value of developed software in relation to the cost and labor consumed while developing that software [43].

As reported by Pressman and Maxim [34], software is designed and developed by the people for the people and maintains a connection between them. People are considered as an important factor in influencing the success or failure of software projects [6]. According to DeMarco and Lister, software projects fail mostly because of incompetent teams [14].

Software teams depend on communication, negotiation, collaboration, and administrative skills to make a project successful [6]. Therefore, appropriate composition of teams is crucial. However, despite the significance of team composition most of the previous studies have focused on technical aspects instead of human, personality, and psychological factors [24].

Only some researchers [6, 14, 24] have investigated the importance of human factors in software development. Their investigations have focused on team members and tasks performed by them to achieve a successful project. Past research [4, 9, 11] has demonstrated that a positive relation exists between personality and team performance.

Personality is defined as attributes and characteristics which make an individual unique [30]. It can be studied by using various popular personality models, such as Myers-Briggs Type Indicator (MBTI) [28], Keirsey Temperament Sorter (KTS) [22], and Five-Factor Model (FFM) also called "Big Five" model [26]. The first two models revolve around personality types while the third is based on personality traits (i.e. human characteristics in diverse dimensions [28]). MBTI, a commonly used

model, employs four components of a character, i.e. Sensing/Instinct, Extraversion/Introversion, Judging/Perceiving, and Thinking/Feeling, while KTS uses four temperaments namely Idealist, Guardian, Artisan, and Rational.

FFM quantifies character using the five personality traits described below [26]:

Openness: An individual with this trait is keen to explore new things and loves creative ideas. In addition, he or she has the power to handle diverse situations in the right manner.

Conscientiousness: Individuals with this trait think and plan everything conscientiously. They prefer to predict the situation and plan accordingly. Once they plan their milestones, they stick to them with great effort and responsibility.

Extraversion: People with this trait are lively, energetic, cheerful, assertive, social, and have extraordinary communication skills.

Agreeableness: People with this trait are trustworthy and have a warm frame of mind. They are always ready to assist others and are of a kind heart.

Neuroticism: Individuals with this trait are inclined to get discouraged, stressed, irritated, and disappointed more frequently.

A number of researchers have conducted qualitative analyses to evaluate the impact of personality on project quality and team productivity [4, 6, 9, 11, 14, 24, 34, 46]. To the best of our knowledge, only two studies have quantitatively measured the team personality and that too by using a measure of central tendency (i.e. mean) which is not always considered a good representative of a dataset. Our previous research [35] was the first study that quantified the notion of team homogeneity using a measure of spread thereby taking data variation and dispersion into account. In that research, we proposed a new metric called Team Homogeneity Index (THI). We also conducted experiments to evaluate the impact of THI on software quality and team productivity during the implementation and testing phases of SDLC.

The aim of this study is to extend our previous research by replicating our experiment with more students in addition to engaging practitioners during software analysis and design, implementation, and testing phases of SDLC to check whether our previous results are generalizable. Furthermore, we have introduced an improved version of THI called Weighted Team Homogeneity Index (WTHI) which uses weights for each of the five personality traits. This improvement is in line with previous research [4, 5, 8, 29] which shows that all traits are not equally important for the software industry. These weights have been obtained by conducting a survey of the Pakistani software industry. Moreover, we have performed a comparative analysis of THI and WTHI with respect to their influence on software quality and team productivity.

Our hypotheses for this research are:

HA0: THI will have no or rather negative relationship with team productivity.

HA1: Teams with higher values of THI will be more productive.

HB0: THI will have no or negative relationship with the quality of software.

HB1: Teams with higher values of THI will produce better quality software.

HC0: WTHI will have no or negative relationship with team productivity.

HC1: Teams with higher values of WTHI will be more productive.

HD0: WTHI will have no or negative relationship with the quality of software.

HD1: Teams with higher values of WTHI will produce better quality software.

HE0: WTHI will be no better in predicting the productivity of teams than THI.

HE1: WTHI will be better in predicting productivity of teams than THI.

HF0: WTHI will be no better in predicting the quality of software than THI.

HF1: WTHI will be better in predicting the quality of software than THI.

The rest of this paper is structured as follows. Section 2 provides a brief summary of related work. The process of determination of weights and calculation of WTHI is described in Section 3. Section 4 describes the assessment criteria while Section 5 provides the details of our experiment. Section 6 discusses the results achieved and Section 7 highlights the threats to the validity of our research. Major conclusions and directions for future work are summarized in Section 8.

2 RELATED WORK

The influence of personality on software teams has been discussed several times in various studies conducted in both industrial and academic environments. In these studies, the main focus was on understanding how much personality influences the performance of a team [4, 7, 9, 11, 13, 24]. The process of assembling a team on the basis of the attributes of its members is known as team composition [25]. These attributes include their experience, demographics, expertise, and other factors regarding their individual personalities [25].

It has been perceived that some individuals can be more productive than others [27, 43]. Similarly, some team members contribute more to the quality of the product as compared to other members. Therefore, it makes a lot of sense to focus on how to create viable teams which can proficiently and effectively develop high quality products [37, 46].

Findings from existing studies show that team composition considerably affects the performance of a team [8]. Some studies recommend that a team should comprise of different personality types to enhance the team's performance [20, 32]. Other studies suggest that a team composed of the same personality types performs better [23, 29].

Different studies have used different personality models to assess the personality of software development teams and have assessed the impact of personality on software quality and team productivity. For example, Rutherfoord [38], Golra and Lam [17], and Sfetsos et al. [42] used KTS as personality assessment tool. Their

results indicated that diverse teams perform better because they communicated and collaborated more as compared to homogeneous teams.

Capretz [10] conducted a survey involving 100 professionals using MBTI personality model. Results revealed that people having diverse skills and personalities form better teams. Peslak [31] also conducted a survey with 55 students. He concluded that extraversion, thinking, and judging personality characteristics positively correlate with project success. Furhtermore, Karn et al. [21], Choi et al. [12], and Poonam and Yasser [33] also used MBTI for personality assessment. Results of [21] revealed that homogeneous teams proved to be highly cohesive and, hence, performed well whereas the results of [12] indicated that the teams with diverse groups were more productive than alike and apposite pair groups. [33] indicated that the performance of pairs working remotely was affected by personality traits.

Walle and Hannay [45] conducted a survey with 88 professionals using FFM personality model. Their results indicated that personality attributes contribute to the collaboration of a team. Later, Salleh et al. [39, 41, 40] used FFM as personality assessment tool in their experiments with 453 students. Their findings showed that Conscientiousness and Neuroticism traits have no significant impact on students' performance whereas a positive correlation was found between openness and teams' performance. Yilmaz et al. [47] also conducted a survey involving 216 professionals. They concluded that practitioners were found to be more extrovert and effective teams were observed to be emotionally more stable.

Acuña et al. [4] calculated team personality by taking the average of personality score of each team member. A positive correlation between the quality of software product and extraversion was found. In 2015, Acuña et al. [5] repeated the same experiment at a larger scale and the results of this experiment were the same as of the first experiment [4]. Furthermore, a positive correlation between product quality and high participative safety and task orientation climate perceptions was observed.

Earlier, we conducted an empirical study [35] to assess the impact of our newly proposed metric, THI, on software quality and team productivity. Our results revealed that THI has a positive impact on different quality factors and team productivity for software implementation and testing phases of SDLC.

3 DETERMINATION OF WEIGHTS AND CALCULATION OF WTHI

As shown in Figure 1, determination of weights is now the first step of our research methodology. To determine the weights of the five personality traits, we conducted a survey [2] of the Pakistani software industry by engaging all members of the Pakistan Software Houses Association (P@sha). A total of 107 professionals belonging to 49 different companies participated in our survey.

Figure 2 shows detailed information about the work experience, rank, and role of these respondents. It also shows the commonly used software development process in the respondent's companies. Figure 2 a) shows that around 50 % of respondents had
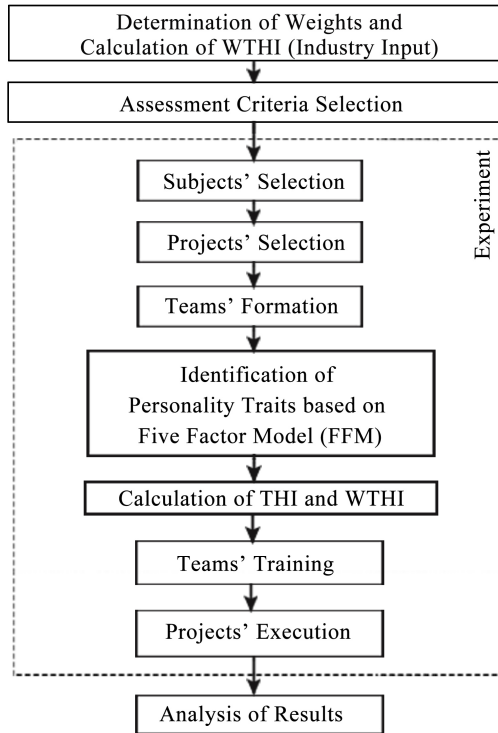
Figure 1. Research methodology (adapted from [35])

four or more years of work experience. This indicates that the input was given by experienced professionals. Figure 2 b) reveals that more than 10 % of the respondents held top management positions like CEO, vice president, and directors in their respective companies. Around 40 % of respondents were playing the role of a team lead or project manager (see Figure 2 c)). It is clear from Figure 2 d), that almost half of the companies represented by these respondents were using Scrum.

Figure 3 shows the frequency of different weights on a scale of 1 to 5 with 1 being least important and 5 being most important provided by respondents for each of the five personality traits. The final value of the weight for each personality trait was calculated by using the arithmetic mean of the weights provided by the respondents. These final values are shown in Figure 4. This figure shows that Openness was considered the most important while Neuroticism was considered the least important personality trait in influencing the software quality and team productivity.

The quantification of THI is a six-step process. The quantification of WTHI also follows the same steps except that it uses weights obtained from the industry (see Figure 4). The complete quantification process is described in Table 1. Figure 5

a) Work experience   b) Rank in company

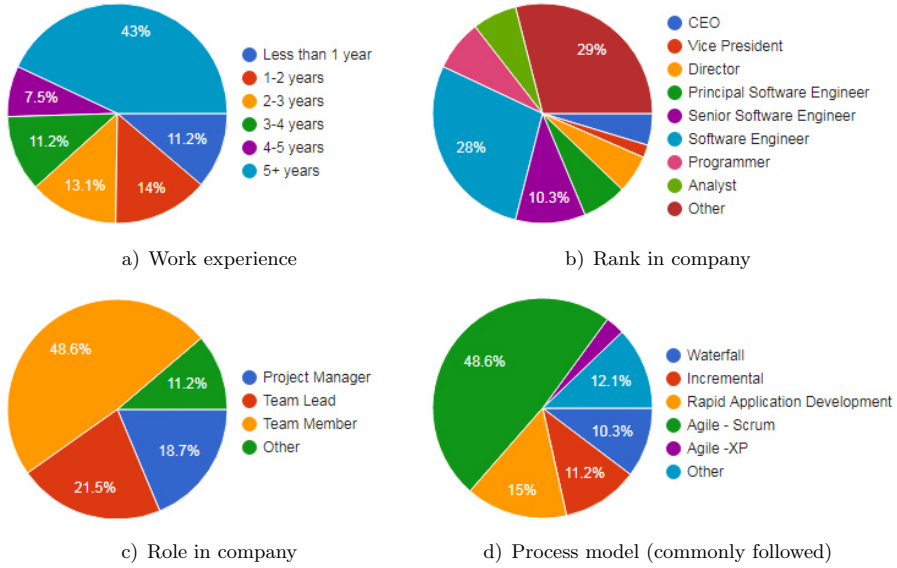c) Role in company   d) Process model (commonly followed)
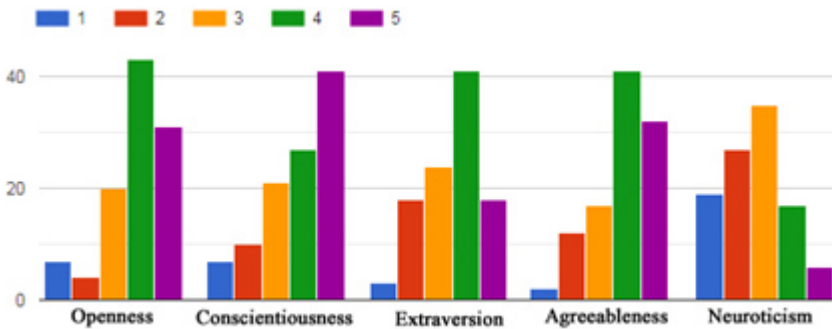
Figure 2. Survey respondents' information

Figure 3. Frequency of weights for each personality trait

shows a detailed worked-out example of calculating WTHI for a five-members team using these steps.

4 ASSESSMENT CRITERIA

4.1 Assessment Criteria for Analysis and Design Phase

Table 2 lists the criteria (adapted from [19]) used to evaluate the quality of the analysis and design models produced by the teams. The quality of these models

Sr.	Step	Description
1	Identification and Normalization of Personality Scores	Identification: 50-item test from IPIP [16] Normalization: Min-Max
2	Calculation of Individual Heterogeneity (find the heterogeneity between the score of one team member and the scores of all other team members)	$$H = |p_k - q_k|$$ where H = Heterogeneity p_k = one team member q_k = other team member
3	Calculation of Overall Heterogeneity (divide the sum of all the weighted heterogeneity by the no. of traits) For WTHI, actual weights (Figure 4) were used For THI, a weight of 1 for all the traits was used	$$OH = 1/n \left(\sum_{k=1}^{n} w_k |p_k - q_k| \right)$$ where OH = Overall Heterogeneity n = total number of personality traits w_k = weight of the kth trait
4	Calculation of Mean (divide the sum of the overall heterogeneity for all the member-pairs with the total number of member-pairs)	$$\text{Mean} = (x_1 + x_2 + x_3 + \cdots + x_n)/m$$ where $x_1, x_2, \ldots, x_n = OH$ of all member-pairs m = all member pairs (i.e. 10)
5	Calculation of Mean Absolute Error (MAE) (subtract the overall heterogeneity of each member-pair from the mean)	$$MAE = 1/m \sum_{k=1}^{m} |\text{mean} - x_k|$$ where MAE = Mean Absolute Error $x_1, x_2, \ldots, x_n = OH$ of all member-pairs m = all member pairs (i.e. 10)
6	Calculation of (W)THI (subtract the MAE from 1) Homogeneity lies between 0 to 1 where 0 indicates no homogeneity and 1 indicates that the team is 100 % homogeneous	$$(W)THI = 1 - MAE$$

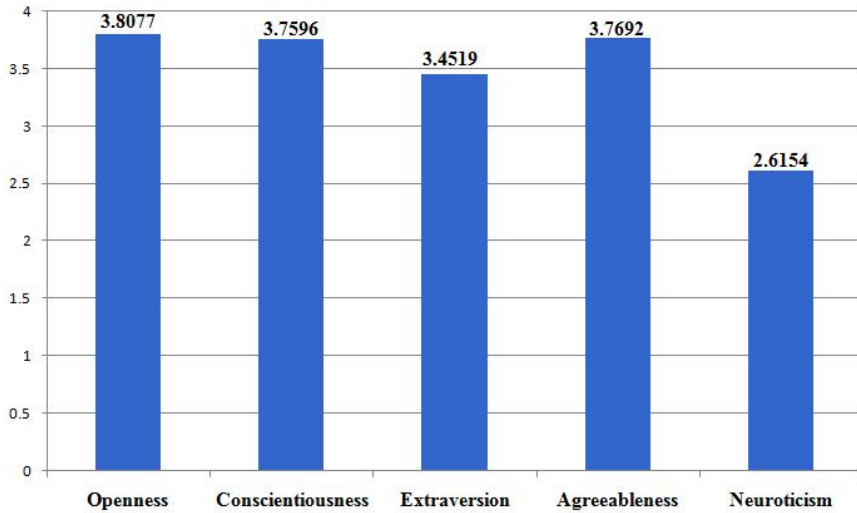Table 1. Quantification process of THI and WTHI

Figure 4. Final weights for each personality trait

was assessed using 24 different factors (grouped under four different categories). The productivity of teams was evaluated using the percentage completeness (using 6 factors) of the analysis and design documents.

4.2 Assessment Criteria for Implementation Phase

Four different quality factors were used to assess the quality of implemented projects i.e. Defect Density, Weighted Sum of Bugs (WSB), Maintainability Index (MI), and Cyclomatic Complexity (CC). The productivity of teams during this phase was determined by taking the ratio of project completeness (weighted sum of implemented features expressed as a percentage and effort taken to complete the project. The detailed assessment criteria for implementation phase are provided in [35].

4.3 Assessment Criteria for Testing Phase

The quality of testing was assessed by looking at Defects Uncovered (number of failed test cases), Architectural Coverage (features tested divided by total features of the project), and Test Case Conformity ((correct test case attributes/total test case attributes) $* 100$) (adapted from [36]). The detailed criteria are given in [35].

5 EXPERIMENT

In order to assess the utility of THI and WTHI in predicting software quality and team productivity, we performed a formal experiment in which software quality and
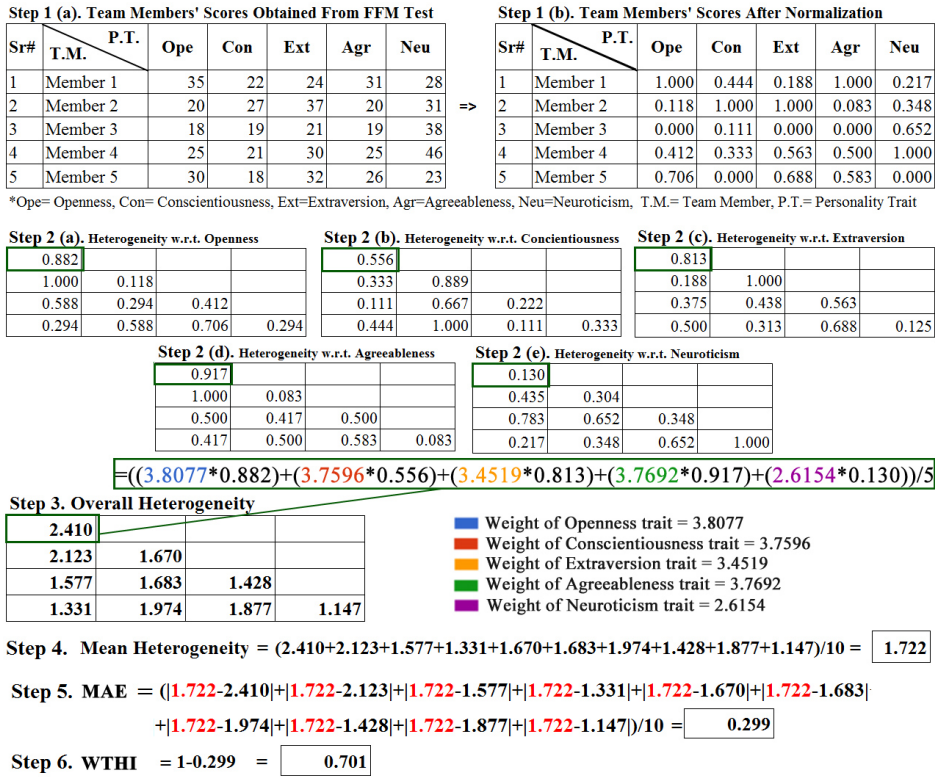
Step 1 (a). Team Members' Scores Obtained From FFM Test

Sr#	T.M. \ P.T.	Ope	Con	Ext	Agr	Neu
1	Member 1	35	22	24	31	28
2	Member 2	20	27	37	20	31
3	Member 3	18	19	21	19	38
4	Member 4	25	21	30	25	46
5	Member 5	30	18	32	26	23

=>

Step 1 (b). Team Members' Scores After Normalization

Sr#	T.M. \ P.T.	Ope	Con	Ext	Agr	Neu
1	Member 1	1.000	0.444	0.188	1.000	0.217
2	Member 2	0.118	1.000	1.000	0.083	0.348
3	Member 3	0.000	0.111	0.000	0.000	0.652
4	Member 4	0.412	0.333	0.563	0.500	1.000
5	Member 5	0.706	0.000	0.688	0.583	0.000

*Ope= Openness, Con= Conscientiousness, Ext=Extraversion, Agr=Agreeableness, Neu=Neuroticism, T.M.= Team Member, P.T.= Personality Trait

Step 2 (a). Heterogeneity w.r.t. Openness

0.882			
1.000	0.118		
0.588	0.294	0.412	
0.294	0.588	0.706	0.294

Step 2 (b). Heterogeneity w.r.t. Concientiousness

0.556			
0.333	0.889		
0.111	0.667	0.222	
0.444	1.000	0.111	0.333

Step 2 (c). Heterogeneity w.r.t. Extraversion

0.813			
0.188	1.000		
0.375	0.438	0.563	
0.500	0.313	0.688	0.125

Step 2 (d). Heterogeneity w.r.t. Agreeableness

0.917			
1.000	0.083		
0.500	0.417	0.500	
0.417	0.500	0.583	0.083

Step 2 (e). Heterogeneity w.r.t. Neuroticism

0.130			
0.435	0.304		
0.783	0.652	0.348	
0.217	0.348	0.652	1.000

$$=((3.8077*0.882)+(3.7596*0.556)+(3.4519*0.813)+(3.7692*0.917)+(2.6154*0.130))/5$$

Step 3. Overall Heterogeneity

2.410			
2.123	1.670		
1.577	1.683	1.428	
1.331	1.974	1.877	1.147

Weight of Openness trait = 3.8077
Weight of Conscientiousness trait = 3.7596
Weight of Extraversion trait = 3.4519
Weight of Agreeableness trait = 3.7692
Weight of Neuroticism trait = 2.6154

Step 4. Mean Heterogeneity = (2.410+2.123+1.577+1.331+1.670+1.683+1.974+1.428+1.877+1.147)/10 = 1.722

Step 5. MAE = (|1.722-2.410|+|1.722-2.123|+|1.722-1.577|+|1.722-1.331|+|1.722-1.670|+|1.722-1.683| +|1.722-1.974|+|1.722-1.428|+|1.722-1.877|+|1.722-1.147|)/10 = 0.299

Step 6. WTHI = 1-0.299 = 0.701

Figure 5. WTHI calculation example (adapted from [35])

team productivity were dependent variables while THI and WTHI were independent variables. Different steps of our experiment (shown in Figure 1) are described below.

5.1 Subjects' Selection

This experiment was carried out in both academic and industrial environments. From the industry, 35 professionals (33 male and 2 female) participated in this experiment while from the academia, a total of 215 BS (Computer Science) students (197 male and 18 female) participated. This experiment was carried out in three different phases of the SDLC, i.e. analysis and design, implementation, and testing. 50 students studying the "Software Engineering" course participated in the analysis and design phase, 90 students enrolled in the "Web Engineering" course participated in the implementation phase, and 75 students registered in the "Software Testing" course took part in the testing phase. The professionals' teams worked on all of the above three phases of SDLC.

Categories	Factors
Completeness	Functionality
	Model Abstraction
	Missing Operations
	Strange Relationships
	Missing Classes and Attributes
	Functions Parameters and their Data Types
Understandability	Complexity
	Easy to Read
	Class, Attributes, and Operations Names
	Relationships Names
	Number of Classes, Operations, and Attributes
	Extra Information
Correctness	Correctness of Data Flow and Layout
	Correctness of Entities
	Correctness of Relationships
	Correctness of Operations
	Correctness of Sequences
	Correctness of Classes
	Correctness of Attributes
	Conformance to the Standards
Layout	Good Class Name
	Neat or Chaotic Structure
	Classes with Similar Size
	Classes Hierarchy and Alignment
	Distance between Classes
	Line Style (Overlapping, Crossing, and Bend)
Use of Relationships at Appropriate Places	Number of Associations
	Number of Aggregations
	Number of Generalizations
	Number of Compositions

Table 2. Analysis and design teams' quality assessment criteria

5.2 Projects' Selection

For professionals, a relatively large project (Online Job Portal) was selected. For students, a comparatively small project (My Shop) was chosen. These same projects were used for all three phases of SDLC.

5.3 Teams' Formation

There were five members in each team of both professionals and students. To keep professionals' teams similar, it was ensured that each team's average experience was between 3–4 years. It was also made sure that educational qualification of every

professional was a bachelors degree in computer science and his/her age was between 24 and 30 years.

To keep students' teams similar, we formed their teams using three buckets of CGPA (i.e. bucket A: 3.00–4.00, bucket B: 2.50–3.00, and bucket C: 2.00–2.50) in such a way that each team consisted of one member from bucket A, two members from bucket B, and two from bucket C. Besides this, it was ensured that no more than one female was assigned to a team and no more than one member in a team had some previous experience related to software development (e.g. internship, freelancing etc.).

5.4 Identification of Personality Traits

Personality traits of all individuals were identified [1] using a 50-item five-factor personality test from the International Personality Item Pool (IPIP) [16].

5.5 Calculation of THI and WTHI

Data obtained in the previous step was used and processed further to compute THI and WTHI. The calculation process of THI and WTHI is explained in detail in [35] and Section 3, respectively.

5.6 Teams' Training

A total of ten training sessions were conducted. There were two types of sessions: a 30-minutes session before the personality assessment test and a 3-hours session before the start of the actual experiment. Two of these ten training sessions were conducted for professionals (one 30-minutes session and one 3-hours training session). Four training sessions were conducted for implementation teams (two 30-minutes sessions and two 3-hours sessions) as there were 90 subjects and to maintain the quality of training sessions they were divided into two groups. Similarly, two sessions (one 30-minutes and one 3-hours) were conducted for each of the analysis and design and testing teams. The agenda of training sessions included explanation of personality test, discussion on Software Requirements Specification (SRS) document provided to teams, guidelines related to experiment, and tutorial of the Time Keeper tool [3] used to record time.

5.7 Projects' Execution

After the successful completion of all the training sessions, the Online Job Portal project was handed over to professionals' teams for performing analysis and design, implementation, and testing. The My Shop project was given to "Software Engineering" students for analysis and design, "Web Engineering" students for implementation, and "Software Testing" students for testing. The maximum time given

to professionals for all three phases combined was 8 weeks (each professional was required to work maximum of 5–6 hours per week). An agreement was reached with the associated software houses that they will have the ownership of the developed projects and we will be allowed to use the project data/documents for research. In the case of students, the time given for analysis and design and implementation of the project was two and four weeks, respectively. Two working days were given for testing.

6 RESULTS AND DISCUSSION

6.1 Analysis and Design Phase

Table 3 shows the details related to the analysis and design models created by the teams. Column 5 contains the scores of models' (class, entity-relationship, data flow, activity, and sequence) overall understandability on a scale of 1 to 10, where 1 means the model is very difficult to understand and 10 means the model is very easy to understand. Overall understandability is calculated by taking the average of the models' individual understandability scores. Column 6 contains the models' overall correctness scores where 1 denotes least correct and 10 represents most correct. Overall correctness is calculated by averaging the models' individual correctness scores. Column 7 contains the models' overall layout scores (obtained by averaging the individual scores). Column 8 contains the average number of relationships (generalization, association, aggregation, and composition) used at appropriate places. Productivity of analysis and design teams appears in the last column.

Figure 6 shows the impact of THI and WTHI on analysis and design teams' models' quality. Figures 6 a) and 6 b) depict the relationship between THI or WTHI and understandability for both students' and professionals' teams. It can be seen that the understandability scores get higher with an increase in the values of THI and WTHI. This supports our hypotheses HB1 and HD1. WTHI seems to have a stronger impact on understandability for students' teams (supporting HF1) but, for professionals' teams, the impact of WTHI on understandability is not relatively strong (HF0 cannot be rejected).

Figures 6 c), 6 d), 6 e) and 6 f) compare the impact of THI and WTHI on students' and professionals' models' correctness scores, layout scores, and appropriately used relationships, respectively. The upward slopes of all trend-lines indicate that correctness, layout and appropriately used relationships have a positive correlation with THI and WTHI supporting our HB1 and HD1 hypotheses. Since WTHI has a stronger impact on correctness, layout, and relationships (as compared to THI), HF1 is also supported for both students' and professionals' teams.

Figure 7 a) displays scatter plots with trend-lines that depict the impact of THI and WTHI on the productivity of students' teams. Figure 7 b) shows the same for professionals' teams. The trend-lines clearly indicate that the teams with greater THI and WTHI values were more productive. This supports our HA1 and HC1 hypotheses. It is worth noting that, for professionals, WTHI has a stronger
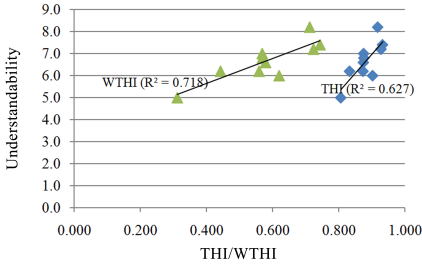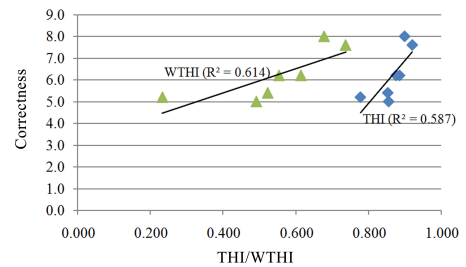
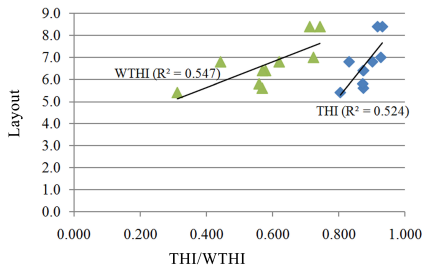a) Impact of THI and WTHI on understandability (students)
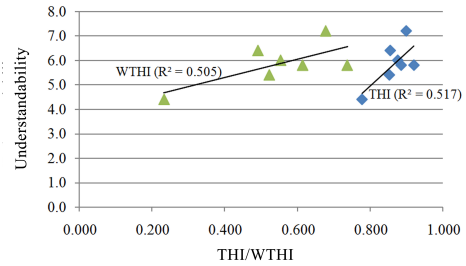
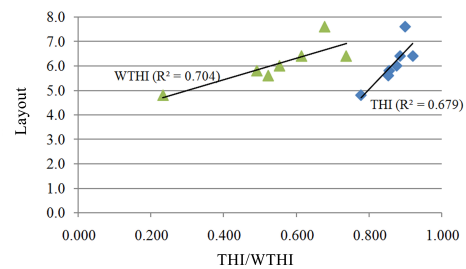b) Impact of THI and WTHI on understandability (professionals)
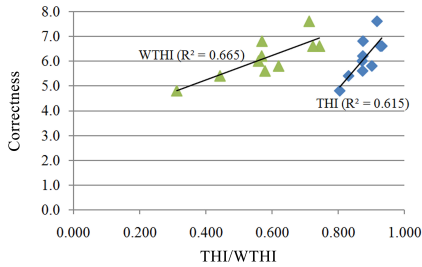
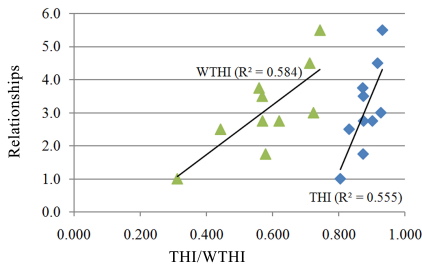c) Impact of THI and WTHI on correctness (students)

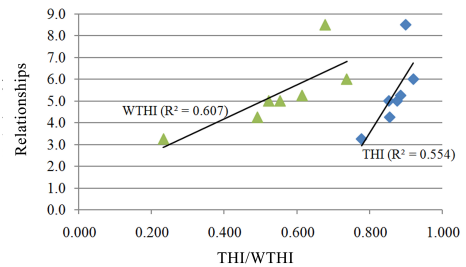d) Impact of THI and WTHI on correctness (professionals)

e) Impact of THI and WTHI on layout (students)

f) Impact of THI and WTHI on layout (professionals)

g) Impact of THI and WTHI on relationship (students)

h) Impact of THI and WTHI on relationship (professionals)

Figure 6. Impact of THI and WTHI on analysis and design teams

Sr.	Teams	THI	WTHI	Und	Cor	Lay	AUR	Prod
1	Stu Team 1	0.874	0.579	6.6	5.6	6.4	1.8	7.487
2	Stu Team 2	0.932	0.743	7.4	6.6	8.4	5.5	10.667
3	Stu Team 3	0.927	0.723	7.2	6.6	7.0	3.0	8.800
4	Stu Team 4	0.902	0.620	6.0	5.8	6.8	2.8	7.111
5	Stu Team 5	0.875	0.569	7.0	6.2	5.6	3.5	7.595
6	Stu Team 6	0.917	0.712	8.2	7.6	8.4	4.5	7.805
7	Stu Team 7	0.805	0.312	5.0	4.8	5.4	1.0	6.593
8	Stu Team 8	0.832	0.443	6.2	5.4	6.8	2.5	6.344
9	Stu Team 9	0.875	0.570	6.8	6.8	6.4	2.8	7.865
10	Stu Team 10	0.873	0.559	6.2	6.0	5.8	3.8	7.564
11	Pro Team 1	0.885	0.614	5.8	6.2	6.4	5.3	2.818
12	Pro Team 2	0.899	0.678	7.2	8.0	7.6	8.5	3.381
13	Pro Team 3	0.777	0.232	4.4	5.2	4.8	3.3	2.625
14	Pro Team 4	0.920	0.737	5.8	7.6	6.4	6.0	3.733
15	Pro Team 5	0.855	0.491	6.4	5.0	5.8	4.3	2.695
16	Pro Team 6	0.876	0.553	6.0	6.2	6.0	5.0	2.895
17	Pro Team 7	0.853	0.522	5.4	5.4	5.6	5.0	2.917
AUR = Appropriately Used Relationships, Cor = Correctness, Lay = Layout, Prod = Productivity, Pro = Professionals, Stu = Students, Und = Understandability								

Table 3. Results of analysis and design phase

positive relationship with productivity. This indicates the contribution of weights assigned and supports the hypothesis HE1 for professionals. In the case of students, hypothesis HE0 cannot be rejected.
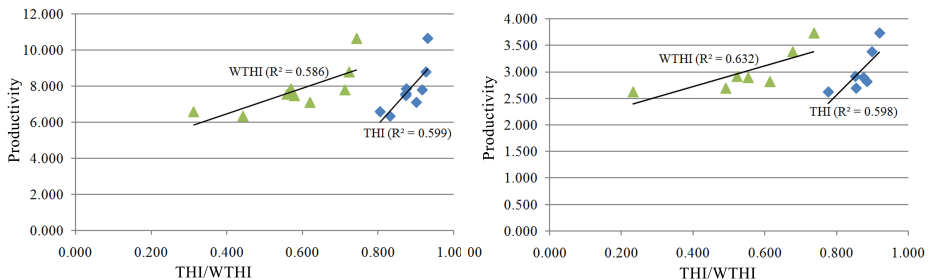
a) Impact of THI and WTHI on productivity (students)
b) Impact of THI and WTHI on productivity (professionals)

Figure 7. Impact of THI and WTHI on analysis and design teams' productivity

6.2 Implementation Phase

The results of the implementation phase also appear favorable. Table 4 provides the data related to the quality of implemented projects and productivity of teams. Weighted sum of bugs (WSB) is provided in column 5 and Defect Density is given in column 6. Maintainability Index (MI) and Cyclomatic Complexity are provided in the last two columns. The last two columns contain the values for productivity and FP productivity of implementation teams.

Sr.	Teams	THI	WTHI	WSB	DD	MI	CC	P	FP
1	Stu Team 1	0.932	0.743	48	0.009	73.330	2.3	0.47	0.96
2	Stu Team 2	0.863	0.501	98	0.023	78.235	1.7	0.37	0.68
3	Stu Team 3	0.906	0.673	64	0.013	61.235	1.5	0.35	0.79
4	Stu Team 4	0.849	0.464	99	0.025	53.255	2.6	0.30	0.58
5	Stu Team 5	0.893	0.620	89	0.028	63.425	1.7	0.32	0.48
6	Stu Team 6	0.891	0.613	70	0.020	67.036	2.5	0.30	0.50
7	Stu Team 7	0.872	0.552	92	0.041	45.651	1.6	0.32	0.30
8	Stu Team 8	0.908	0.675	58	0.016	43.392	2.8	0.40	0.52
9	Stu Team 9	0.833	0.426	140	0.057	60.424	3.9	0.25	0.35
10	Stu Team 10	0.889	0.609	97	0.043	47.235	2.3	0.38	0.34
11	Stu Team 11	0.863	0.508	93	0.043	64.936	2.6	0.36	0.36
12	Stu Team 12	0.895	0.649	91	0.026	55.456	2.2	0.37	0.67
13	Stu Team 13	0.900	0.696	85	0.026	76.253	1.9	0.33	0.46
14	Stu Team 14	0.866	0.511	91	0.026	61.436	1.6	0.33	0.50
15	Stu Team 15	0.919	0.655	87	0.023	77.219	1.5	0.29	0.56
16	Stu Team 16	0.847	0.440	114	0.075	57.945	2.5	0.26	0.24
17	Stu Team 17	0.932	0.782	91	0.016	79.548	2.3	0.41	0.85
18	Stu Team 18	0.785	0.232	146	0.062	52.235	4.5	0.21	0.337
19	Pro Team 1	0.885	0.614	131	0.016	72.050	4.8	0.31	0.95
20	Pro Team 2	0.899	0.678	123	0.013	78.820	4	0.43	0.95
21	Pro Team 3	0.777	0.232	156	0.021	66.040	5.6	0.28	0.74
22	Pro Team 4	0.920	0.737	106	0.012	85.670	2.6	0.42	0.95
23	Pro Team 5	0.855	0.491	173	0.020	77.890	4.9	0.28	0.92
24	Pro Team 6	0.876	0.553	133	0.016	69.390	4.7	0.29	0.90
25	Pro Team 7	0.853	0.522	138	0.016	76.650	4.9	0.29	0.91
CC = Cyclomatic Complexity, DD = Defect Density, Pro = Professionals, MI = Maintainability Index, Stu = Students, WSB = Weighted Sum of Bugs, P = Productivity, FP = Function Point Productivity									

Table 4. Results of implementation phase

Figure 8 shows the impact of THI and WTHI on the quality of implemented projects. Figures 8 a), 8 b), 8 c) and 8 d) depict the impact on the weighted sum of bugs and defect density of implemented projects. The descending slopes of trend-lines clearly support HB1 and HD1 and indicate that the teams with higher THI

and WTHI values developed software with lower weighted bugs and defect density. It is also evident from these figures that HF1 is supported for professionals only.

The effects of THI and WTHI on maintainability index and cyclomatic complexity are shown in Figures 8 e), 8 f), 8 g) and 8 h). The upward sloping trend-lines for maintainability index and downward slopping trend-lines for cyclomatic complexity indicate that the teams with greater THI and WTHI values developed projects with greater maintainability and lower complexity. Hence, these figures support our HB1 and HD1 hypotheses. HF1, again, is supported for professionals only.

Figure 9 shows scatter plots with trend-lines that depict the relationship of THI and WTHI with productivity and FP productivity. These upward sloping trend-lines for productivity and FP productivity indicate that teams with higher values of THI and WTHI are more productive. These findings support our HA1 and HC1 hypotheses. In the case of professionals' teams only, R^2 values for WTHI are greater than those for THI for both FP productivity and productivity. Thus, hypothesis HE1 is supported for professionals only.

6.3 Testing Phase

The details regarding the testing are presented in Table 5. Figure 10 shows the impact of THI and WTHI on quality of testing. It is clear from these upward sloping trend-lines in Figures 10 a) and 10 b) that THI and WTHI have a positive correlation with architectural coverage. Hence, hypotheses HB1 and HD1 are supported for both students' and professionals' teams. HF1 is supported for professionals' teams only.

Figures 10 c), 10 d), 10 e) and 10 f) show scatter plots with associated trend-lines that display the impact of THI and WTHI on total number of defects uncovered by testing teams and conformity to provided test case template. Clearly, teams with higher values of THI and WTHI uncovered more defects and followed the given test case template more strictly. Hence, HB1 and HD1 are supported. HF1 is also supported for both types of teams.

Figure 11 shows the relationships between THI and productivity and WTHI and productivity of testing teams. It is clear from this figure that the teams with greater THI and WTHI values appear more productive. These values support our HA1 and HC1 hypotheses. In the case of professionals' teams only, the R^2 values for WTHI are greater than those for THI. This shows that hypothesis HE1 is supported for professionals' teams only.

Our research focuses on the impact of team homogeneity (i.e. WTHI) on team productivity and software quality. Our results imply that human factors (i.e. personality aspects or team homogeneity) should be taken into consideration while assigning jobs to existing employees or hiring new personnel. WTHI can help the software industry managers during the team composition process. Our results indicate that teams with higher values of THI and WTHI are more productive and produced better quality software. Our results also reveal that WTHI is more strongly correlated with software quality and team productivity for professionals' teams in comparison with THI. The results of the industry survey conducted to determine
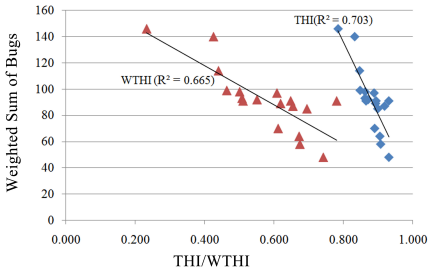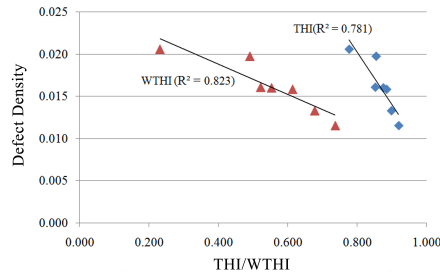
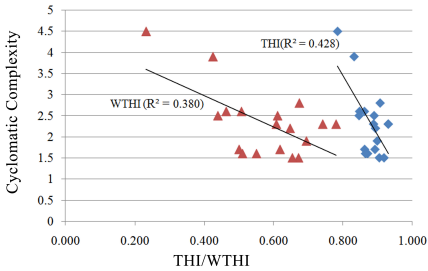a) Impact of THI and WTHI on weighted sum of bugs (students)
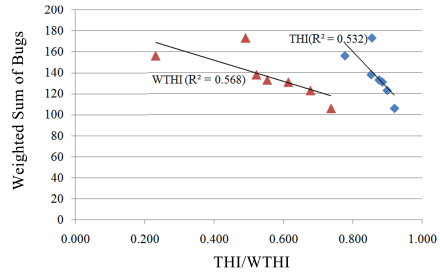
b) Impact of THI and WTHI on weighted sum of bugs (professionals)
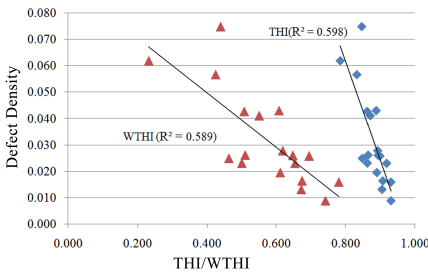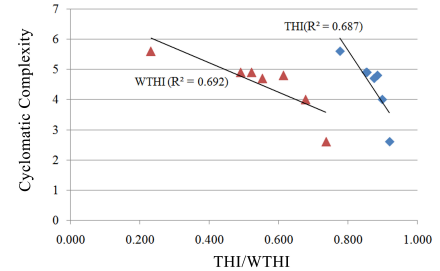
c) Impact of THI and WTHI on defect density (students)
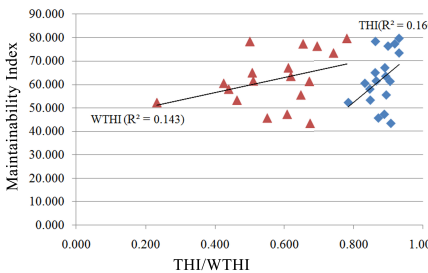
d) Impact of THI and WTHI on defect density (professionals)
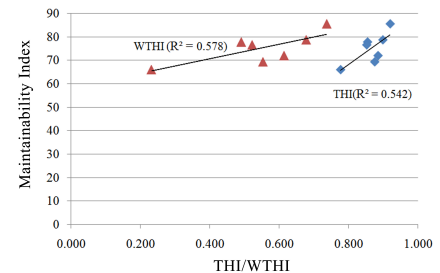
e) Impact of THI and WTHI on cyclomatic complexity (students)

f) Impact of THI and WTHI on cyclomatic complexity (professionals)

g) Impact of THI and WTHI on maintainability index (students)

h) Impact of THI and WTHI on maintainability index (professionals)

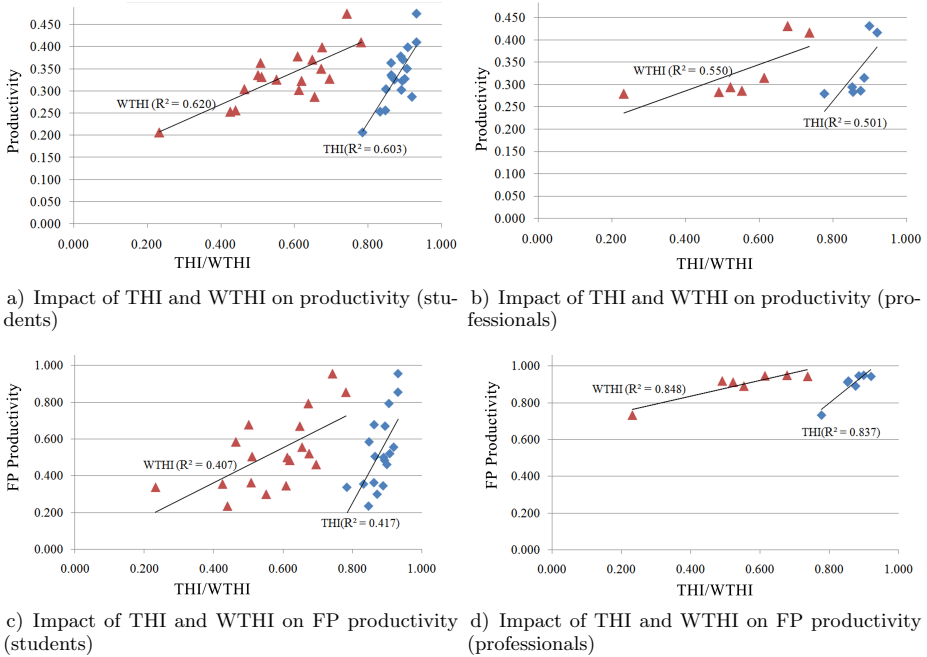Figure 8. Impact of THI and WTHI on implemented projects' quality

a) Impact of THI and WTHI on productivity (students)

b) Impact of THI and WTHI on productivity (professionals)

c) Impact of THI and WTHI on FP productivity (students)

d) Impact of THI and WTHI on FP productivity (professionals)

Figure 9. Impact of THI and WTHI on productivity of implementation teams

the weights of personality traits show that "Openness to Experience", "Agreeableness", and "Conscientiousness" are the top three most important traits whereas the "Neuroticism" is the least important trait.

6.4 Hypotheses Testing

We have investigated our null hypotheses based on one-way analysis of variance (ANOVA) test [18] to analyze the significant difference between independent and dependent variables for students' and professionals' teams. THI and WTHI are our independent variables whereas different components of team productivity and software quality are our dependent variables. A threshold p-value of 0.05 was used.

Table 6 provides the details of the 60 hypothesis tests (30 for THI and 30 for WTHI) we have conducted using the IBM SPSS tool [44] for different factors of software quality and team productivity. It can be seen that in the case of productivity during the analysis and design phase, the null hypothesis (HA0) can be rejected for both students' and professionals' teams. For almost all of the quality factors considered for analysis and design models, hypotheses (HB1 and HD1) can be accepted. The only exception is understandability (for both THI and WTHI in case of professionals). Similarly, for team productivity during the implementation

| Sr.	Teams	THI	WTHI	AC	DU	TCC	Prod
1	Stu Team 1	0.808	0.328	37.255	5	83.239	1.508
2	Stu Team 2	0.850	0.444	50.980	6	84.244	1.898
3	Stu Team 3	0.905	0.680	60.784	8	91.364	2.952
4	Stu Team 4	0.873	0.577	35.294	6	86.898	1.552
5	Stu Team 5	0.827	0.386	27.451	4	85.795	0.950
6	Stu Team 6	0.883	0.585	50.980	4	90.385	2.122
7	Stu Team 7	0.799	0.312	19.608	2	80.519	0.654
8	Stu Team 8	0.825	0.396	29.412	5	87.190	1.136
9	Stu Team 9	0.896	0.664	47.059	7	92.375	1.691
10	Stu Team 10	0.807	0.337	23.529	3	88.462	1.371
11	Stu Team 11	0.903	0.650	23.529	5	89.744	2.200
12	Stu Team 12	0.888	0.611	39.216	6	87.500	2.139
13	Stu Team 13	0.862	0.507	33.333	5	90.476	2.061
14	Stu Team 14	0.856	0.478	33.333	4	84.921	1.789
15	Stu Team 15	0.878	0.563	37.255	5	90.705	1.818
16	Pro Team 1	0.885	0.614	111.765	12	91.390	1.606
17	Pro Team 2	0.899	0.678	131.373	16	93.006	1.775
18	Pro Team 3	0.777	0.232	70.588	11	88.060	1.059
19	Pro Team 4	0.920	0.737	101.961	17	94.921	2.306
20	Pro Team 5	0.855	0.491	82.353	11	87.619	1.331
21	Pro Team 6	0.876	0.553	111.765	12	93.704	1.755
22	Pro Team 7	0.853	0.522	100.000	3	91.870	1.709
AC = Architectural Coverage , DU = Defects Uncovered , Pro = Professionals, Prod = Productivity, TCC = Test Case Conformity, Stu = Students							

Table 5. Results of testing phase

phase null hypothesis (HA0) is rejected for productivity and FP productivity in all cases except for the productivity of professionals' teams (for THI).

In case of quality factors during the implementation phase, we cannot reject the null hypothesis HB0 only for the weighted sum of bugs (for THI in case of professionals) and maintainability index (for both THI and WTHI). For the testing phase, the results are more promising for both productivity and quality. The null hypothesis cannot be rejected only for THI in the case of professionals' teams' defects uncovered. Out of 60 tests, we are unable to reject the null hypotheses for just 8 cases. Furthermore, in 5 out of those 8 cases overall, null hypotheses are rejected for only professionals' teams for THI. This shows the importance of using weights especially for professionals' teams.

7 THREATS TO VALIDITY

Despite the fact that our results seem promising, some factors may threaten their validity. First of all, in the case of students, subjects' competence, intelligence,
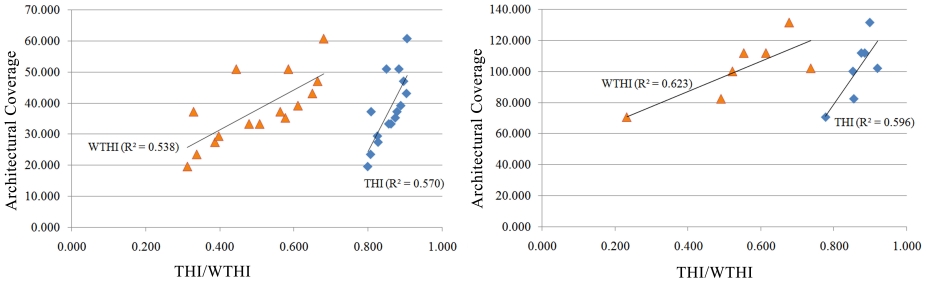
a) Impact of THI and WTHI on architectural coverage (students)

b) Impact of THI and WTHI on architectural coverage (professionals)
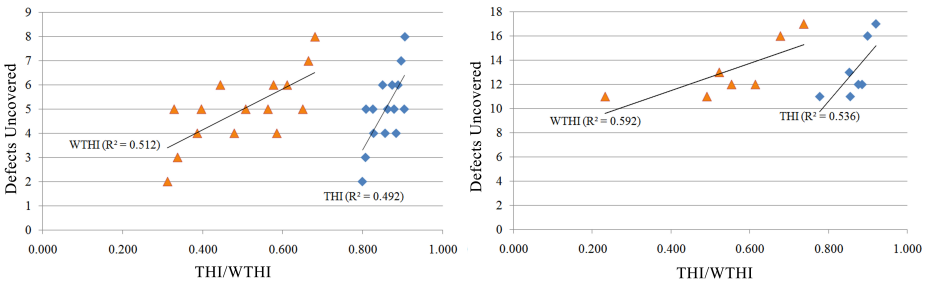
c) Impact of THI and WTHI on defects uncovered (students)

d) Impact of THI and WTHI on defects uncovered (professionals)
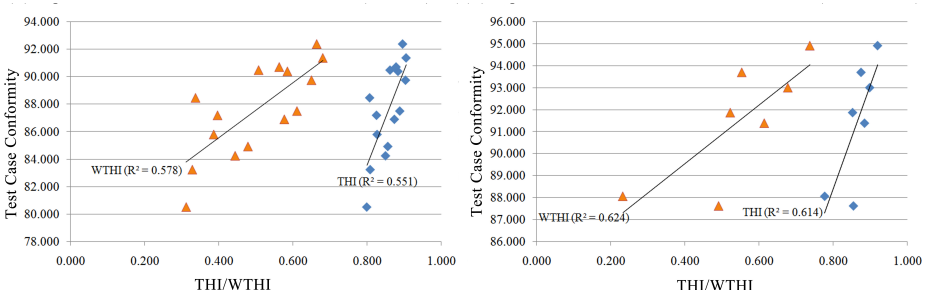
e) Impact of THI and WTHI on test case (students)

f) Impact of THI and WTHI on test case (professionals)

Figure 10. Impact of THI and WTHI on testing teams' test cases' quality

learning ability, degree of friendship, previous domain knowledge, programming experience, interest in programming, and gender can influence team's productivity and quality of projects. To avoid these threats, we formulated teams by making three buckets of CGPA (I. 3 or above, II. 2.50 to 3.00, III. 2.00 to 2.50) and randomly selected one member from bucket I and two members each from buckets II and III. At most one member with some previous experience was part of a single team. Also, there was no more than one female member in each team. Moreover,
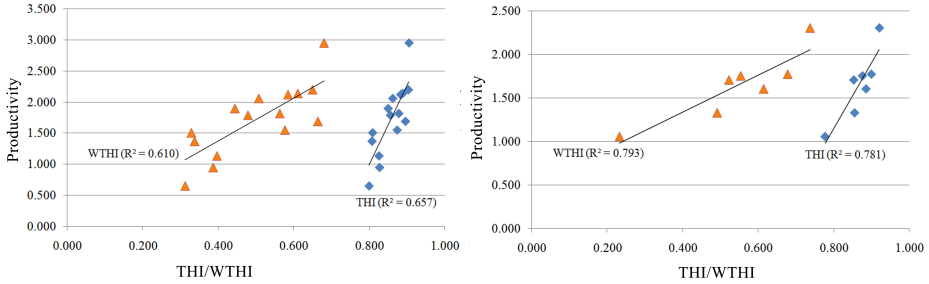
a) Impact of THI and WTHI on productivity (students)

b) Impact of THI and WTHI on productivity (professionals)

Figure 11. Impact of THI and WTHI on productivity of testing teams

it was made sure that the average CGPA of teams was within a specific range (i.e. 2.50 to 2.80).

In the case of professionals, level of experience, academic qualifications, and domain exposure can affect team productivity and project quality. These threats were mitigated by making sure that every team member has a Bachelors degree in Computer Science. Furthermore, the average experience of a team was kept between 3 to 4 years and no more than one female member was assigned to a team.

Last, but not the least, the selection of different projects for each iteration of a specific phase of SDLC could have made our results incomparable. We avoided this threat by selecting the same project for all iterations of the experiment for each phase of SDLC. Moreover, using the same project for professionals as well as students could have compromised team productivity and project quality. We eliminated this threat by selecting a relatively more complex project for professionals. This helped us in avoiding threats to external validity thereby making our results generalizable to both academic and industrial environments.

8 CONCLUSIONS AND FUTURE WORK

In this research we introduced a new metric called WTHI (derived from THI). A survey was conducted in the Pakistani software industry and weights were determined for all five personality factors. The impact of THI and WTHI was compared for analysis and design, implementation, and testing phases of SDLC in academic and industrial environments.

The results of this study reveal that teams with greater values of THI and WTHI performed better in almost all the phases of SDLC. A positive correlation of team productivity and software quality was observed with both THI and WTHI. These results also indicate that, as compared to THI, WTHI is more strongly correlated with team productivity and software quality for all the productivity and quality factors in the case of teams comprising professionals.

Sr.	Dependent Variables	Sub	F	THI (Sig)	F	WTHI (Sig)
Analysis and Design Phase						
1	Productivity	Stu	11.99	0.01	11.34	0.01
2		Pro	7.45	0.04	8.62	0.03
3	Understandability	Stu	13.45	0.01	20.43	0.00
4		Pro	5.37	0.07	5.11	0.07
5	Correctness	Stu	12.79	0.01	15.90	0.00
6		Pro	7.11	0.04	7.96	0.04
7	Layout	Stu	8.83	0.02	9.69	0.01
8		Pro	10.58	0.02	11.92	0.02
9	Relationships	Stu	10.00	0.01	11.25	0.01
10		Pro	5.53	0.05	6.83	0.04
Implementation Phase						
11	Productivity	Stu	24.33	0.00	26.21	0.00
12		Pro	5.02	0.08	6.12	0.05
13	FP Productivity	Stu	11.48	0.00	11.01	0.00
14		Pro	25.69	0.00	28.05	0.00
15	Weighted Sum of Bugs	Stu	38.01	0.00	31.79	0.00
16		Pro	5.68	0.06	6.58	0.05
17	Defect Density	Stu	23.80	0.00	23.00	0.00
18		Pro	17.86	0.01	23.36	0.00
19	Cyclomatic Complexity	Stu	11.97	0.00	9.84	0.01
20		Pro	11.02	0.02	11.27	0.02
21	Maintainability Index	Stu	3.06	0.10	2.68	0.12
22		Pro	5.92	0.06	6.68	0.05
Testing Phase						
23	Productivity	Stu	24.92	0.00	20.39	0.00
24		Pro	17.87	0.01	19.23	0.01
25	Architectural Coverage	Stu	17.25	0.00	15.18	0.00
26		Pro	7.39	0.04	8.27	0.03
27	Defects Uncovered	Stu	12.61	0.00	13.65	0.00
28		Pro	5.79	0.06	7.26	0.04
29	Test Case confirmity	Stu	15.98	0.00	17.82	0.00
30		Pro	7.98	0.04	8.32	0.03
Sig. = Significance of F or p-value, Sub = Subject, Pro = Pro, Stu = Students						

Table 6. Results of ANOVA tests

Future work in this direction may focus on evaluating the impact of WTHI on other phases of SDLC such as requirements engineering and maintenance. It would also be interesting to replicate this experiment using more complex projects and larger team sizes (i.e. more than five members in a team). Last, but not the least, other personality models (e.g. MBTI, KTS, etc.) may also be used to determine team homogeneity.

REFERENCES

[1] Personality Test Based on Five Factor Model.

[2] A Study on Personality Traits' Rating.

[3] Time Keeper.

[4] ACUÑA, S. T.—GÓMEZ, M.—JURISTO, N.: How Do Personality, Team Processes and Task Characteristics Relate to Job Satisfaction and Software Quality? Information and Software Technology, Vol. 51, 2009, No. 3, pp. 627–639, doi: 10.1016/j.infsof.2008.08.006.

[5] ACUÑA, S. T.—GÓMEZ, M. N.—HANNAY, J. E.—JURISTO, N.—PFAHL, D.: Are Team Personality and Climate Related to Satisfaction and Software Quality? Aggregating Results from a Twice Replicated Experiment. Information and Software Technology, Vol. 57, 2015, pp. 141–156, doi: 10.1016/j.infsof.2014.09.002.

[6] ACUÑA, S. T.—JURISTO, N.: Assigning People to Roles in Software Projects. Software: Practice and Experience, Vol. 34, 2004, No. 7, pp. 675–696, doi: 10.1002/spe.586.

[7] BARROSO, A. S.—MADUREIRA, J. S.—SOARES, M. S.—DO NASCIMENTO, R. P. C.: Influence of Human Personality in Software Engineering – A Systematic Literature Review. Proceedings of the 19th International Conference on Enterprise Information Systems (ICEIS), 2017, Vol. 1, pp. 53–62, doi: 10.5220/0006292000530062.

[8] BELL, S. T.: Deep-Level Composition Variables as Predictors of Team Performance: A Meta-Analysis. Journal of Applied Psychology, Vol. 92, 2007, No. 3, pp. 595–612, doi: 10.1037/0021-9010.92.3.595.

[9] BRADLEY, J. H.—HEBERT, F. J.: The Effect of Personality Type on Team Performance. Journal of Management Development, Vol. 16, 1997, No. 5, pp. 337–353, doi: 10.1108/02621719710174525.

[10] CAPRETZ, L. F.: Personality Types in Software Engineering. International Journal of Human-Computer Studies, Vol. 58, 2003, No. 2, pp. 207–214, doi: 10.1016/s1071-5819(02)00137-4.

[11] CAPRETZ, L. F.—AHMED, F.: Making Sense of Software Development and Personality Types. IT Professional, Vol. 12, 2010, No. 1, pp. 6–13, doi: 10.1109/mitp.2010.33.

[12] CHOI, K. S.—DEEK, F. P.—IM, I.: Exploring the Underlying Aspects of Pair Programming: The Impact of Personality. Information and Software Technology, Vol. 50, 2000, No. 11, pp. 1114–1126, doi: 10.1016/j.infsof.2007.11.002.

[13] CRUZ, S.—DA SILVA, F. Q. B.—CAPRETZ, L. F.: Forty Years of Research on Personality in Software Engineering: A Mapping Study. Computers in Human Behavior, Vol. 46, 2015, pp. 94–113, doi: 10.1016/j.chb.2014.12.008.

[14] DEMARCO, T.—LISTER, T.: Peopleware: Productive Projects and Teams. Third Edition. Addison-Wesley, 2013.

[15] GALIN, D.: Software Quality Assurance: From Theory to Implementation. Pearson Education India, 2004.

[16] GOLDBERG, L. R.—JOHNSON, J. A.—EBER, H. W.—HOGAN, R.—ASHTON, M. C.—CLONINGER, C. R.—GOUGH, H. G.: The International Personality Item Pool and the Future of Public-Domain Personality Mea-

sures. Journal of Research in Personality, Vol. 40, 2006, No. 1, pp. 84–96, doi: 10.1016/j.jrp.2005.08.007.

[17] GORLA, N.—LAM, Y. W.: Who Should Work with Whom? Building Effective Software Project Teams. Communications of the ACM, Vol. 47, 2004, No. 6, pp. 79–82, doi: 10.1145/990680.990684.

[18] JAMES, G.—WITTEN, D.—HASTIE, T.—TIBSHIRANI, R.: An Introduction to Statistical Learning. Springer Texts in Statistics Book Series, Vol. 103, 2013, doi: 10.1007/978-1-4614-7138-7.

[19] KARASNEH, B.—STIKKOLORUM, D.—LARIOS, E.—CHAUDRON, M.: Quality Assessment of UML Class Diagrams. Proceedings of the MODELS Educators Symposium 2015, Ottawa, Canada, 2015. CEUR Workshop Proceedings, Vol. 1555, 2015.

[20] KARN, J.—COWLING, T.: A Follow Up Study of the Effect of Personality on the Performance of Software Engineering Teams. Proceedings of the 2006 ACM/IEEE International Symposium on Empirical Software Engineering (ISESE '06), 2006, pp. 232–241, doi: 10.1145/1159733.1159769.

[21] KARN, J. S.—SYED-ABDULLAH, S.—COWLING, A. J.—HOLCOMBE, M.: A Study Into the Effects of Personality Type and Methodology on Cohesion in Software Engineering Teams. Behaviour and Information Technology, Vol. 26, 2007, No. 2, pp. 99–111, doi: 10.1080/01449290500102110.

[22] KEIRSEY, D.—BATES, M.: Please Understand Me: Character and Temperament Types. Prometheus Nemesis Books, Del Mar, CA, 1978.

[23] KICHUK, S. L.—WIESNER, W. H.: The Big Five Personality Factors and Team Performance: Implications for Selecting Successful Product Design Teams. Journal of Engineering and Technology Management, Vol. 14, 1997, No. 3-4, pp. 195–221, doi: 10.1016/s0923-4748(97)00010-6.

[24] LENBERG, P.—FELDT, R.—WALLGREN, L. G.: Behavioral Software Engineering: A Definition and Systematic Literature Review. Journal of Systems and Software, Vol. 107, 2015, pp. 15–37, doi: 10.1016/j.jss.2015.04.084.

[25] LEVINE, J. M.—MORELAND, R. L.: Progress in Small Group Research. Annual Review of Psychology, Vol. 41, 1990, No. 1, pp. 585–634, doi: 10.1146/annurev.ps.41.020190.003101.

[26] MCCRAE, R. R.—JOHN, O. P.: An Introduction to the Five-Factor Model and Its Applications. Journal of Personality, Vol. 60, 1992, No. 2, pp. 175–215, doi: 10.1111/j.1467-6494.1992.tb00970.x.

[27] MEYER, A. N.—FRITZ, T.—MURPHY, G. C.—ZIMMERMANN, T.: Software Developers' Perceptions of Productivity. Proceedings of the 22nd ACM SIGSOFT International Symposium on Foundations of Software Engineering (FSE 2014), 2014, pp. 19–29, doi: 10.1145/2635868.2635892.

[28] MYERS, I. B.—MCCAULLEY, M. H.—QUENK, N. L.—HAMMER, A. L.: MBTI Manual: A Guide to the Development and Use of the Myers-Briggs Type Indicator. Third Edition. Consulting Psychologists Press, Palo Alto, CA, 1998.

[29] PEETERS, M. A.—VAN TUIJL, H. F.—RUTTE, C. G.—REYMEN, I. M.: Personality and Team Performance: A Meta-Analysis. European Journal of Personality, Vol. 20, 2006, No. 5, pp. 377–396, doi: 10.1002/per.588.

[30] Pervin, L. A.—John, O. P.: Handbook of Personality: Theory and Research. Second Edition. Elsevier, 1999.

[31] Peslak, A. R.: The Impact of Personality on Information Technology Team Projects. Proceedings of the 2006 ACM SIGMIS CPR Conference on Computer Personnel Research: Forty Four Years of Computer Personnel Research: Achievements, Challenges and the Future (SIGMIS CPR '06), 2006, pp. 273–279, doi: 10.1145/1125170.1125233.

[32] Pieterse, V.—Kourie, D. G.—Sonnekus, I. P.: Software Engineering Team Diversity and Performance. Proceedings of the 2006 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on IT Research in Developing Countries (SAICSIT '06), 2006, pp. 180–186, doi: 10.1145/1216262.1216282.

[33] Poonam, R.—Yasser, C. M.: An Experimental Study to Investigate Personality Traits on Pair Programming Efficiency in Extreme Programming. 2018 5th International Conference on Industrial Engineering and Applications (ICIEA), IEEE, 2018, pp. 95–99, doi: 10.1109/iea.2018.8387077.

[34] Pressman, R. S.: Software Engineering: A Practitioner's Approach. Sixth Edition. Palgrave Macmillan, 2005.

[35] Qamar, N.—Malik, A. A.: Birds of a Feather Gel Together: Impact of Team Homogeneity on Software Quality and Team Productivity. IEEE Access, Vol. 7, 2019, pp. 96827–96840, doi: 10.1109/access.2019.2929152.

[36] Qamar, N.—Malik, A. A.: Evaluating the Impact of Pair Testing on Team Productivity and Test Case Quality – A Controlled Experiment. Pakistan Journal of Engineering and Applied Sciences, Vol. 25, 2019, pp. 80–88.

[37] Richardson, I.—Casey, V.—McCaffery, F.—Burton, J.—Beecham, S.: A Process Framework for Global Software Engineering Teams. Information and Software Technology, Vol. 54, 2012, No. 11, pp. 1175–1191, doi: 10.1016/j.infsof.2012.05.002.

[38] Rutherfoord, R. H.: Using Personality Inventories to Help Form Teams for Software Engineering Class Projects. ACM SIGCSE Bulletin, Vol. 33, 2001, No. 3, pp. 73–76, doi: 10.1145/507758.377486.

[39] Salleh, N.—Mendes, E.—Grundy, J.: The Effects of Openness to Experience on Pair Programming in a Higher Education Context. 2011 24th IEEE-CS Conference on Software Engineering Education and Training (CSEE & T), IEEE, 2011, pp. 149–158, doi: 10.1109/cseet.2011.5876082.

[40] Salleh, N.—Mendes, E.—Grundy, J.—Burch, G. S. J.: The Effects of Neuroticism on Pair Programming: An Empirical Study in the Higher Education Context. Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM '10), 2010, Art. No. 22, 10 pp., doi: 10.1145/1852786.1852816.

[41] Salleh, N.—Mendes, E.—Grundy, J.—Burch, G. S. J.: An Empirical Study of the Effects of Conscientiousness in Pair Programming Using the Five-Factor Personality Model. Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering (ICSE '10), 2010, Vol. 1, pp. 577–586, doi: 10.1145/1806799.1806883.

[42] Sfetsos, P.—Stamelos, I.—Angelis, L.—Deligiannis, I.: An Experimental Investigation of Personality Types Impact on Pair Effectiveness in Pair Programming. Empirical Software Engineering, Vol. 14, 2009, No. 2, Art. No. 187, doi: 10.1007/s10664-008-9093-5.

[43] Sudhakar, G. P.—Farooq, A.—Patnaik, S.: Measuring Productivity of Software Development Teams. Serbian Journal of Management, Vol. 7, 2012, No. 1, pp. 65–75, doi: 10.5937/sjm1201065s.

[44] IBM SPSS Statistics.

[45] Walle, T.—Hannay, J. E.: Personality and the Nature of Collaboration in Pair Programming. 2009 3rd International Symposium on Empirical Software Engineering and Measurement, Lake Buena Vista, FL, USA, IEEE, 2009, pp. 203–213, doi: 10.1109/esem.2009.5315996.

[46] Yilmaz, M.—O'Connor, R. V.—Clarke, P.: Effective Social Productivity Measurements During Software Development – An Empirical Study. International Journal of Software Engineering and Knowledge Engineering, Vol. 26, 2016, No. 3, pp. 457–490, doi: 10.1142/s0218194016500194.

[47] Yilmaz, M.—O'Connor, R. V.—Colomo-Palacios, R.—Clarke, P.: An Examination of Personality Traits and How They Impact on Software Development Teams. Information and Software Technology, Vol. 86, 2017, pp. 101–122, doi: 10.1016/j.infsof.2017.01.005.
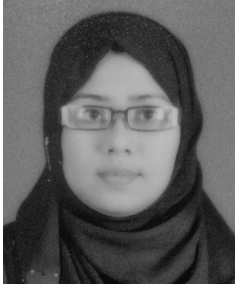
Nosheen Qamar is currently working as Assistant Professor at the Department of Computer Science and Information Technology, the University of Lahore, Lahore, Pakistan. She received her Ph.D. degree in computer science from the National University of Computer and Emerging Sciences (FAST-NUCES), Lahore, Pakistan in 2020. Before joining the University of Lahore, she has gained 7 years of industry experience with last job title as Software Development Manager. Her research interests include software engineering, design patterns, software teams, project management, empirical software engineering, and requirements engineering.

Ali Afzal Malik is currently working as Assistant Professor and Head of the Computer Science Department of the National University of Computer and Emerging Sciences (FAST-NUCES). He started his professional career in 2003 working as Software Engineer in Techlogix – a well-reputed Pakistani software house. After receiving the prestigious Fulbright scholarship in 2005, he obtained M.S. and Ph.D. degrees in computer science from the University of Southern California (USC), Los Angeles, USA in 2007 and 2010, respectively. He was awarded the Office of International Services (OIS) Academic Achievement Award twice (2007 and 2010) during his stay at USC. His research paper on the quantitative aspects of requirements elaboration was given the best paper award in SBES 2008 (Sao Paulo, Brazil). Before joining FAST-NUCES in 2013, he has held an adjunct faculty position at the Lahore University of Management Sciences (LUMS) and a full-time faculty position at the University of Central Punjab (UCP). He has undertaken research in software cost estimation at two of the world's leading research centers in software engineering i.e. USC's Center for Systems and Software Engineering (CSSE) and Institute of Software, Chinese Academy of Sciences (ISCAS). His current research work focuses on areas such as empirical software engineering, requirements engineering, and software cost estimation. He is a senior member of IEEE.