An Empirical Study of the Effects of Personality in Pair Programming using the Five-Factor Model

Norsaremah Salleh
Department of Computer Science
University of Auckland
Auckland, New Zealand
nsal017@aucklanduni.ac.nz

Emilia Mendes
Department of Computer Science
University of Auckland
Auckland, New Zealand
emilia@cs.auckland.ac.nz

John Grundy
Dept. of Electrical and Computer Engineering
University of Auckland
Auckland, New Zealand
John-g@cs.auckland.ac.nz

Giles St. J Burch
Management and International Business
University of Auckland
Auckland, New Zealand
g.burch@auckland.ac.nz

Abstract— **Pair Programming (PP) has been long researched in industry and academia. Although research evidence about its usefulness is somewhat inconclusive, previous studies showed that its use in an academic environment can benefit students in programming and design courses. In our study, we investigated the "human" aspect of PP; in particular the effects that personality attributes may have on PP's effectiveness as a pedagogical tool. We conducted a formal experiment at the University of Auckland to investigate the influence of personality differences among paired students using the Five-Factor Model as a personality measurement framework. The aim of our study was to improve the implementation of PP as a pedagogical tool through understanding the impact the variation in the personality profile of paired students has towards their academic performance. Our findings showed that differences in personality traits did not significantly affect the academic performance of students who pair programmed. In addition, the majority (88%) of students were satisfied with the PP experience and 87% responded that their confidence level increased when pairing.**

Keywords-component; Pair programming; experimentation; personality type; five-factor model

I. INTRODUCTION

The adoption of pair programming (PP) as a pedagogical tool in the Computer Science/Software Engineering (CS/SE) curriculum has received significant attention among researchers [21], [30], [31]. As one of the Extreme Programming practices, PP was reported to be beneficial for students not only for introductory programming courses, but also in other CS/SE courses [25].

When using PP, one student acts as the "driver" i.e. as someone who is playing the key role in handling the tasks using a computer. Meanwhile the partner (the "navigator") observes the driver's work and provides support by giving ideas or pointing to errors in their code or design [56]. Pairs swap roles after a given time so that each partner can experience both roles. It was reported by Sfetsos et al. that such interaction between the driver and the navigator was critical in determining the level of PP's success [55]. In their investigation, they found that the XP system was mainly driven by "people" factors. Since the collaborative work demonstrated by PP naturally involves a psychosocial interaction between two people working together, it seems imperative to investigate PP's psychosocial elements [12].

Based on evidence from our systematic review of PP in higher education, we found that personality was the most common factor investigated in previous PP studies. However, in terms of the effect or influence of personality towards PP's effectiveness, existing results were inconsistent [51]. Research evidence also suggests that developers' personality is one of PP's most critical success factors [4], [13], [33]. Therefore, the aim of our study was to improve the implementation of the PP practice as a pedagogical tool by focusing on personality traits and demonstrate evidence any improvements. More specifically, we assessed whether differences in students' personality profile when pairing did actually impact upon their academic performance. Our study contributes to the PP body of knowledge by providing evidence in understanding the role of personality in determining the successful implementation of PP and consequently benefiting educators in terms of providing suggestions to implement pair formation effectively. Another contribution of our study is to ameliorate the learning outcome of CS/SE students in higher education institutions.

This paper reports on a formal experiment conducted with undergraduate students at The University of Auckland during the 2009 summer school. The remainder of this paper is organized as follows: Section 2 describes our motivation and related work. Section 3 briefly introduces the Five Factor Model as basis for the personality profiling used in our study. Section 4 presents the research methodology followed by the presentation of results in Section 5. Finally we draw our conclusions in Section 6.

II. MOTIVATION AND RELATED WORK

Research in CS/SE typically involves a human element as one of its important aspects of investigation. However, it has been reported that there has been too much focus on the techniques, processes, and methods involved in developing software, neglecting the human issues [23]. Feldt et al. [23] suggested that software engineering (SE) empirical studies should embark on gathering psychometric data on the people involved in software development. In particular, their study focused on understanding the role of personality towards the attitude to SE tools and processes. Their findings showed that an individual's personality traits correlated with their

attitude towards work style, and adaptability to changes. Since PP is a practice that involves people working together to achieve a common set of goals, the practice's success is largely determined by how effective they work as a team, despite their skills or abilities.

In the context of team effectiveness, personality has been reported to be a critical success factor in determining teamwork success among students [2],[39],[36]. In this regard, numerous studies have investigated SE team performance and effective team composition based on personality traits [51],[52]. Pieterse & Kourie [52] looked at the role of personality within teams of tertiary students. They found out that the diversity of personalities in a team had significant positive impact on a team's success. In another study conducted by Katira et al., results were mixed regarding the effect of personality towards compatibility of paired students [37]. In their study, MBTI was used for measuring personality.

Table 1 summarizes the existing PP studies conducted in academic and industrial settings that investigated the impact of personality traits in PP. In general, their findings were quite diverse and most studies reported that personality had no significant influence in determining PP's effectiveness [11], [32], [38], [26]. We believe that numerous factors may have contributed to obtaining these diverse results, such as different experimental settings, which include the level of studies (undergraduate or postgraduate), nature of courses, instructors, and the instruments used to measure personality.

TABLE I. SUMMARY OF LITERATURE RELATING PERSONALITY TO PP

Author(s)	Type of study	Sub.	Size	IV	DV	Outcomes	Personality measurement
Chao & Atli [11]	Survey & Exp.	Stud.	58	Personality traits	PP success (code quality and pair compatibility)	PP success is not influenced by differences in personality traits.	Personality characteristics (Univ. of Denver Career Centre)
Heiberg et al. [32]	Formal Exp.	Stud.	110	PP Vs. Non-PP	PP productivity	The individual personality traits do not have significant consequences to PP performance.	NEO PI
Katira et al. [37]	Formal Exp.	Stud.	564	Personality, skill level, technical competence, and self-esteem	Pair compatibility	Results were mixed. Personality differences affect compatibility of freshmen but not for advanced undergraduate students	MBTI
Katira et al. [38]	Formal Exp.	Stud.	361	Personality, skill level, self esteem, gender and ethnicity	Pair compatibility	Pair compatibility was not affected by personality of the paired students.	MBTI
Layman [40]	Survey	Stud.	119	All paired	Perception towards collaboration	Personality had no significant effect towards perception to collaborate.	MBTI
Sfetsos et al. [54]	Formal Exp.	Stud.	84	Personality	PP effectiveness	Paired of mixed personalities performed better than paired of same personality.	Keirsey Temperament Sorter
Williams et al. [57]	Formal Exp.	Stud.	1350	Personality, learning style, skills, self esteem, work ethic.	Pair compatibility	Results were mixed. Partial supports of personality in predicting compatibility.	MBTI
Choi K.S. [12]	Formal Exp.	Stud.	128	Personality	PP outcome (code productivity)	Personality differences have significant impact on PP outcomes.	MBTI
Gevaert [26]	Formal Exp.	Stud.	28	PP Vs Solo	Time spent	Personality does not significantly affect the efficiency of students who paired	Eysenck Personality Questionnaire
Dick & Zarnett [18]	Case studies	Prof.	8	N/A	N/A	Personality traits critical for PP success were communication, comfortableness working in a team, confidence and ability to compromise.	N/A
Exp – Experiment Sub. – Subject Stud – Student Unk. – Unknown N/A – Not available Prof. – Professional IV – Independent Variable DV – Dependent Variable							

Most PP studies investigated personality type using the Myers-Briggs Type Indicator (MBTI) as a personality assessment method [37],[38],[40],[57],[12]. MBTI is one of the most popular instruments used to measure an individual's personality based on four basic dimensions: Extroversion vs. Introversion, Sensing vs. Intuition, Thinking vs. Feeling, and Judging vs. Perceiving [48]. In the area of consultancy and training, MBTI is commonly used as a personality measure [24]. However, psychology researchers investigating personality usually apply the

five-factor NEO Personality Inventory (NEO-PI) [15],[22]. The NEO-PI was constructed based on the Five Factor Model (FFM) of personality, which is currently considered the predominant taxonomy of personality by personality psychologists, and consists of "big-five" traits: Agreeableness, Conscientiousness, Extraversion, Neuroticism, and Openness to experience [8].

To date, Heiberg et al. [32] was the only PP study that applied NEO-PI to measure personality type (see Table 1). They suggested that an individual programmer's personality had no significant effect on PP's effectiveness. However, they did not investigate PP's effectiveness based on a combination of personalities in a single pair. Such a clear lack of evidence motivated us to investigate personality's effect towards PP by applying the FFM theoretical framework. The selection of personality traits as variables would provide an advantage in overcoming the problem of bad pairing experience reported in some PP studies [40],[34]. The discomfort or incompatibility experienced working with a partner might be due to a mismatch in psychosocial aspects such as personality, and gender combinations. Cockburn & Williams highlighted that understanding the social aspects of PP is critical towards attaining success of the practice [14]. This is mainly because the PP practice is a collaborative process involving interaction and communication between two people working together to achieve a common set of goals. As different people possess different ideas and opinions, understanding how the two students can work best together is imperative to the success of PP as a pedagogical tool.

In our formal experiment, personality was measured using an online publicly accessible set of items for personality assessment known as "International Personality Item Pool Representation of the NEO-PI-R" (IPIP-NEO) [27],[28]. The selection of IPIP-NEO as our personality assessment was due to two major reasons: (1) It is based on the FFM framework, and (2) It provides a Web interface for collecting and scoring calculation of personality responses, which is much more efficient compared with the paper-based version [7],[29]. The personality scales in IPIP-NEO are also represented numerically, thus allowing us to perform more powerful statistical methods when compared with the "binary" type of MBTI scales [23]. In the following Section, we introduce the FFM theoretical framework, which is the basis of the personality assessment used in our study.

III. FIVE-FACTOR MODEL

The Five-Factor Model characterises five broad personality traits - Openness to experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism – and provides a structure that categorizes dimensions of differences in human personality [43]. Openness to experience describes intellectual, cultural, or creative interest. Conscientiousness is concerned with one's achievement orientation. People who are high on conscientiousness tend to be hardworking, organized, able to complete tasks thoroughly, and also reliable, whereas

low conscientiousness relates to negative traits such as being irresponsible, impulsive, and disordered. Extraversion relates to the degree of sociability, gregariousness, assertiveness, talkativeness, and activeness [1]. Agreeableness refers to positive traits such as cooperativeness; kindness, trust and warmth, and persons low on agreeableness tend to be sceptical, selfish, and hostile. Neuroticism refers to the state of emotional stability. Someone low in neuroticism tends to appear calm, confident, and secure, whereas a highly neurotic individual tends to be moody, anxious, nervous, and insecure [19].

In comparison with MBTI, FFM was derived based on research on the classic trait theory, whereas MBTI was developed based on Jung's theory of psychological type [24]. In terms of the scoring method used to measure personality, MBTI classifies an individual's personality into 1 of 16 different types using the combination of the four dichotomous preferences. In FFM, the scoring is made by summing the scores of each facet from each factor, using a five-point likert scale. Thus, MBTI uses a bipolar discontinuous scale, in contrast to a continuous scale used by the FFM.

Our selection of FFM as a personality assessment framework was due to its comprehensive nature and its ability to capture the basic temperament and dispositional factors relevant to the educational context [17]. In terms of its validity and reliability, FFM is generally accepted by personality psychologists who suggest that such a broad trait of dimensions adequately represents human personality attributes [1],[2].

IV. RESEARCH METHODOLOGY

This Section details the formal experiment conducted during 2009 summer school at the University of Auckland.

A. Research Objectives

Our experiment's research objectives were outlined using the Goal/Question Metric (GQM) framework [3]. The GQM definition is shown in Table 2, and the detailed goal definition for the formal experiment is the following:

Object of study: PP technique.
Purpose: To improve the effectiveness of PP as a pedagogical tool in higher education institutions.
Focus: To investigate the influence of personality as a psychosocial factor that can potentially affect the success of the PP practice in CS/SE courses/tasks.
Perspective: From the point of view of the researcher
Context: In the context of undergraduate CS/SE students.

B. Context

The formal experiment was conducted during the 2009 summer school involving first year students enrolled in an introductory programming course. Students paired during compulsory weekly tutorials, run by a tutor and a few TAs.

Goal(s)	Question(s)	Metric(s)
To investigate the effect of personality differences towards successful pair configuration	Do differences in personality type affect PP's effectiveness?	Students' academic achievement measured by assignments and test scores
To investigate the level of satisfaction and confidence of paired students.	Were students satisfied and did they feel confidence working in pairs?	PP questionnaire on satisfaction and confidence level

C. Research Goals

Our research aims to improve the effectiveness of PP as a pedagogical tool for CS/SE education by investigating the effects that personality differences among paired students may have on PP's effectiveness. The primary purpose of the investigation was to increase students' satisfaction and amount of learning.

D. Hypotheses

Existing literature suggests that diversity or heterogeneity of personality among team members is a strong predictor of team success [36],[52],[6]. In a follow up study of the effect of personality on the performance of SE teams, Karn & Cowling reported that a team consisting of members of heterogeneous personality worked well together [36]. Similar findings were also documented in [9],[52]. Their studies however, were conducted in the context of teams consisting of four to five members. In our study we focused on the same issues but involving peer or pair collaboration. In order to investigate the effect of personality differences on PP's effectiveness, we proposed the following null hypothesis:

H_O: Differences in personality traits do not affect the effectiveness of students who pair programmed.

which is contrasted by the following alternative hypothesis:

H_A: Differences in personality traits affect the effectiveness of students who pair programmed.

Differences in personality can be operationalized by forming pairs consisting of students with different levels of conscientiousness. Previous findings showed conscientiousness to consistently positively predict educational success [17],[9], [20]. Thus, in our study we also chose conscientiousness as the main personality factor believed to affect PP's effectiveness. High conscientiousness is always related to being a high achiever, organized, and thorough, whereas low conscientiousness possesses the opposite traits such as low need for achievement, being unprepared and disorganized [43]. Table 4 shows the categorization of pairs according

to personality differences using as basis the conscientiousness factor. Pair (C_{High}, C_{High}) denotes that a pair consists of students with very similar personality (higher scores on conscientiousness). Meanwhile, Pair (C_{High}, C_{Low}) refers to pairs of very different personality (higher and lower scores on conscientiousness).

Similar Personality	Mixed Personality
Pair (C_{Low}, C_{Low})	Pair (C_{Low}, C_{Med})
Pair (C_{Med}, C_{Med})	Pair (C_{Med}, C_{High})
Pair (C_{High}, C_{High})	Pair (C_{Low}, C_{High})

We hypothesized that pairs consisting of mixed personalities would achieve better academic performance compared with pairs of students with similar personalities. Our experiment also looked into the association between students' personality scores and their academic achievement, level of satisfaction and confidence.

E. Variables

Our synthesis of evidence from the systematic review showed that measuring PP's effectiveness could be achieved using "academic performance", "technical productivity", "program quality", or "satisfaction". Since our study aimed at facilitating CS/SE students through the practice of PP, the metrics to measure PP's effectiveness were "academic performance" and students' "satisfaction". Hence, personality traits were our independent variables, and PP's effectiveness and satisfaction our dependent variables. PP's effectiveness was measured using assignments and test scores, and satisfaction was measured using a questionnaire where all questions employed a nine-point likert-scale.

F. Instrumentation and Materials

During the first course lecture, one of the authors gave an overview of the experiment (including PP) and distributed consent forms and participant information sheets (PIS) to the students for signing. The PIS described important information regarding the experiment and highlighted its major purpose.

We used a short version of the IPIP-NEO to measure participants' personality traits. Although the original version of IPIP-NEO provided a more reliable result, the short version of IPIP-NEO was reported to measure exactly the same traits and to also present acceptable measurement reliability [35]. The IPIP-NEO[1] consists of 120 items which descriptions were authored by John A. Johnson [35]. The test produces scores in a numerical scale, with 0 being the lowest score, and 99 the highest score for each factor. Based on the suggestion described in [35], the personality traits were classified into low, average or high based on the range of scores shown in Table 4.

---

[1] The test can be accessed at this public domain URL: http://www.personal.psu.edu/ /j5j/IPIP/

TABLE IV.   PERSONALITY SCORES LEVEL

Scores	Lowest 30%	Middle 40%	Highest 30%
Level	Low	Average	High

In addition to the online test, participants were administered with a pre-test questionnaire to gather their demographic information as well as their programming competency level. At the end of each tutorial (about 10 minutes before the tutorial dispersed), participants were given a short questionnaire to measure their level of satisfaction working with their partner. Inline with the University's requirement, we have sought the approval by the University of Auckland's Human Participants Ethics Committee prior to performing the data collection.

The formal experiment was carried out involving undergraduate computer science (CS) students enrolled in an introductory programming course (COMPSCI101). During the course, students learnt about basic programming concept in Java, and created a few small applications as part of their assignments. The rationale for using students as subjects was mainly due to the study's focus - to improve PP's effectiveness in an academic setting.

G. Experimental Procedure

Each of the tutorials was treated as an independent experiment. Our hypothesis was investigated using a "single factor between-group design" [47] as the experimental design. This design allows each subject to experience only one condition or group, which means, in a particular tutorial, a student was assigned either to a pair of similar personality or to a pair of mixed personality (*controlled group* = similar personality, *experimental group* = mixed personality). Therefore, before the first tutorial (i.e. during the first week of semester), students' personality data were gathered using the online IPIP test. The results of the personality test were used to allocate partners. For this purpose, the personality scores of conscientiousness were used to assign students between two different groups of similar or mixed personality (e.g. A student with higher score on conscientiousness was paired with someone with low score on conscientiousness to form a pair of mixed personality).

Every tutorial lasted for two hours. During this time, the tutor explained a topic for about 45 minutes, followed by exercises for the remaining 75 minutes. To allow for "pair-jelling", students worked with their partners for an initial period of 30 minutes; and then swapped their roles every 15-20 minutes. Before the end of every tutorial, students provided feedback working with the partner by filling out a questionnaire. The exercises given during the tutorials were graded, thus contributing towards the students' final grade. In addition, assignments and test were also graded but completed individually.

The outcomes measured from the experiment were the students' academic performance in their test and three assignments. Since tutorials varied from week to week, the experiments were designed in such a way to minimize the confounding factor which might occur due to differences in tasks and level of complexity of exercises assigned to the students. Therefore, the tasks and exercises remained the same throughout the week.

H. Analysis Procedure

In order to test our null hypothesis, we used a single factor multivariate analysis of variance (MANOVA) to analyse the difference in academic performance between the controlled and experimental groups. We used a Mann-Whitney U statistic to measure the ordinal variable "satisfaction" against our independent variable. In terms of measuring the association between variables, we used the bivariate Pearson correlation. The statistical package to generate the results of our analysis was SPSS v. 17.

V. RESULTS

In this Section, the results from the formal experiment are presented, followed by a discussion of results and summary of threats to the validity of our findings.

A. Subjects

The subjects involved in the formal experiment were 54 undergraduate CS students. Sixty-five percent (65%) of the subjects were male, and subjects' age ranged from 19 to 30 years (median = 20). Of the 32 students who responded to the demographic survey, 84.4% indicated that they did not have any work experience. Two students dropped out from the course, thus, they were excluded from our analysis. Of 52 students, only 49 students completed the personality test.

B. Correlations between Big Five Traits and Academic Performance

The distribution of test scores between pairs of same and mixed personality can be seen from the boxplot shown in Figure 1. The scores between the two groups were somewhat similar, but on average, paired students from a mixed personality group obtained higher marks than their counterparts.
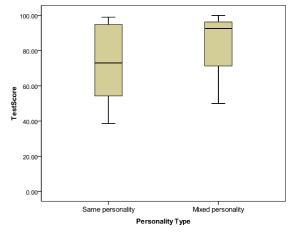

Figure 1.   Comparison of test scores between personality groups

In terms of individual achievement in students' assignments, there was no significant difference between the two groups (see Figure 2). Students obtained higher marks regardless of the personality differences in their pairing experience. Notice that there were some outliers in the boxplot. These outliers represented students who did not submit some of the assignments, thus affecting their overall assignments' performance.
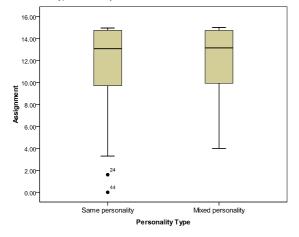

Figure 2. Comparison of assignments scores between personailty group

Figure 3 shows three boxplots of scores for each level of conscientiousness. The distributions of scores between the boxplots have a similar spread, but the median scores for students of low conscientiousness outperformed the other two groups (average and high). We noticed that some of the students from this group (i.e. low conscientiousness) have had several years of working experience and reported to have greater programming competency than their peers.
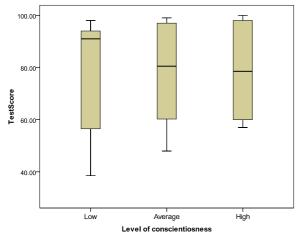

Figure 3. Comparison of test scores between conscientiousness level

In assessing the relationship between variables, one can measure the strength of a relationship using a correlation test [47]. Table 5 provides the matrix correlation between the five personality factors and students' academic performance (test and assignments scores). Conscientiousness and openness to experience were the two traits that showed positive correlation with students' performance, but the results were mixed. Conscientiousness showed a positive association with assignments' scores (r=0.266), but no correlation with test scores. This result indicates that highly conscientious students typically scored higher marks for their assignments regardless of their pairing configuration.

The only personality factor that had a significant correlation with test scores was the openness to experience (r^2=0.12 and r is 0.35). This finding corroborates that of another study [22] which reported that openness to experience was positively correlated with undergraduate academic success.

TABLE V. CORRELATION BETWEEN THE ACADEMIC PERFORMANCE AND PERSONALITY FACTORS (N=46)

	1	2	3	4	5	6	7
1	1	0.36**	-0.01	0.075	0.27*	-0.16	0.15
2		1	0.08	0.19	0.07	-0.04	0.35*
3			1	0.07	0.30**	-0.49**	0.32*
4				1	0.18	-0.07	0.12
5					1	-0.53*	-0.02
6						1	0.06
7							1

1. Assignments 2. Test 3. Extraversion 4. Agreeableness 5. Conscientiousness 6. Neuroticism 7. Openness to Experience ** Correlation is significant at the 0.01 (1-tailed) * Correlation is significant at the 0.05 (1-tailed)

C. HypothesesTtesting

We used a single factor multivariate analysis of variance (MANOVA) to analyze whether there was any significant difference in academic achievement between similar and mixed personality groups of paired students. MANOVA is regarded as a complex statistic that linearly combines several dependent variables in a single analysis, where variables need to be correlated at a low to moderate level [41]. Herein, assignments and test scores were analyzed simultaneously using the General Linear Model program in SPSS.

Table 6 provides mean values and standard deviation values for assignments and test scores, for each group. Mean differences are almost similar for assignments' scores but somewhat different for test scores. Table 7 shows the results for differences on performance between the two groups. MANOVA generated four multivariate tests (by default). Of these four tests, the one that provides "good and commonly used multivariate F" is Wilks' Lambda [41]. Thus, referring to Wilks' Lambda (under the "PairType" effect), results showed no significant difference (F=2.513, df=43, p=0.93) between the "PairType" groups, on a linear combination of two dependent variables (assignments and test scores). Thus, using the 95% confidence interval we failed to reject the null hypothesis based on our data, thus supporting the view that personality traits did not affect the effectiveness of students who pair programmed.

TABLE VI. MEAN AND STANDARD DEVIATION OF PAIRED STUDENTS OF SIMILAR AND MIXED PERSONALITY

	Personality Type	Mean	SD	N
Assignments	Same personality	12.30	3.62	23
	Mixed personality	11.99	3.02	23
	Total	12.15	3.3	46
Test Scores	Same personality	75.57	21.33	23
	Mixed personality	83.52	16.39	23
	Total	78.04	19.61	46

TABLE VII. MULTIVARIATE TESTS

Effect		Value	F	Hyp. Df	Error df	Sig.
Intercept	Pillai's Trace	0.96	490	2.0	43.0	0.0
	Wilk's Lambda	0.04	490	2.0	43.0	0.0
	Hotelling's Trace	22.79	490	2.0	43.0	0.0
PairType	Roy's Largest Root	0.11	2.51	2.0	43.0	0.09
	Wilk's Lambda	0.89	2.51	2.0	43.0	0.09
	Hotelling's Trace	0.12	2.51	2.0	43.0	0.09
	Roy's Largest Root	0.12	2.51	2.0	43.0	0.09

D. Results on Satisfaction and Confidence

The response rate of the post-experimental survey was approximately 67% in every tutorial. The surveys were distributed in the second week of the semester until the final week of tutorials (altogether nine tutorials). Data were analyzed separately as each tutorial was treated as a single independent "mini-experiment".

Our analysis showed that overall students obtained a large amount of satisfaction and confidence from the pairing activity (see Figures 4 and 5). In terms of satisfaction, on average 88.54% students were satisfied working with their partner, and approximately 87.88% responded that their level of confidence solving the exercises with their partner was high. On average, 92.6% students enjoyed working collaboratively with their partner.
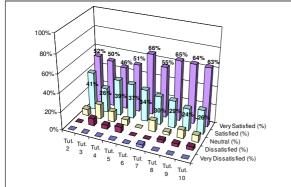
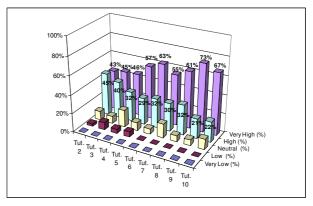Figure 4. Survey on PP Satisfaction

Figure 5. Survey on PP Confidence

To answer the question on whether there are any differences in satisfaction level between the controlled and experimental groups, we applied the Mann-Whitney U test to each of the experiments' unit. Nonparametric testing was chosen because the dependent variable (i.e. satisfaction level) was not normally distributed thus violating the assumptions for parametric testing. In Table 8, the mean satisfaction ranks for paired students are shown. The group with the highest mean rank had the highest level of satisfaction.

Although the similar personality group appeared to score higher ranks in most of the experiments, these differences were not always significant. As can be seen in Table 9, using a significance level of 0.05, there were no significant differences between groups. Overall, results demonstrated that the satisfaction levels of paired students were not affected by personality differences, and paired students achieved higher satisfaction regardless of their differences in personality when pairing.

TABLE VIII. MANN-WHITNEY U RANKS FOR SATISFACTION LEVEL

		N	Mean Rank	Sum of Ranks
Tut. 2 (N=39)	Same Personality	13	23.46	305.0
	Mixed Personality	26	18.27	475.0
Tut. 3 (N=37)	Same Personality	23	21.09	485.0
	Mixed Personality	14	15.57	218.0
Tut. 4 (N=36)	Same Personality	22	17.55	386.0
	Mixed Personality	14	20.0	280.0
Tut. 5 (N=26)	Same Personality	15	14.43	216.5
	Mixed Personality	11	12.23	134.5
Tut. 6 (N=34)	Same Personality	9	18.17	163.5
	Mixed Personality	25	17.26	431.5
Tut. 7 (N=31)	Same Personality	10	16.3	163.0
	Mixed Personality	21	15.86	333.0
Tut. 8 (N=30)	Same Personality	13	18.31	238.0
	Mixed Personality	17	13.35	227.0
Tut. 9 (N=31)	Same Personality	15	17.6	264.0
	Mixed Personality	16	14.5	232.0
Tut. 10 (N=24)	Same Personality	13	15.5	201.5
	Mixed Personality	11	8.95	98.5

TABLE IX.   MANN-WHITNEY U TEST STATISTICS FOR
SATISFACTION LEVEL

	Mann-Whitney U	Wilcoxon W	Z	Asymp. Sig. (2-tailed)	Exact Sig.
Tut. 2	124.0	475.0	-1.51	0.13	0.19
Tut. 3	113.0	218.0	-1.62	0.11	0.14
Tut. 4	133.0	386.0	-0.75	0.45	0.51
Tut. 5	68.5	134.5	-0.79	0.43	0.47
Tut. 6	106.5	431.5	-0.28	0.78	0.82
Tut. 7	102.0	333.0	-0.14	0.89	0.92
Tut. 8	74.0	227.0	-1.80	0.72	0.13
Tut. 9	96.0	232.0	-1.07	0.29	0.36
Tut. 10	32.5	98.5	-2.63	0.01	0.02

E. Discussion

We observed an interesting finding with regard to relating paired students' performance with personality traits. We found that the personality traits that appeared to have a positive correlation with academic performance were conscientiousness and openness to experience. This result is in line with several existing studies in psychology [10],[22],[15], and business [49]. They found that conscientious students were more likely to perform well in the class compared with low conscientious students.

Our failure to support the alternative hypotheses can be attributed to several reasons. In terms of sample size, we believe that a larger sample size is needed because the ability to detect differences across the studied groups is greater in a larger sample size [16]. There is also an issue with regard to the process of forming a pair of similar or mixed personality. For example, matching a high conscientious with a low conscientious student can possibly produce an incompatible pair due to dissimilarities in character and attitude (Based on the survey, we noted comments from students in mixed personality pairs who did not enjoy working with their partner). Likewise, forming a pair of similar personality where both are low conscientious students may bring disadvantages to the pair due to their lack of self-discipline and low need for achievement. In this sense, comparing the performance of paired students between similar and mixed personalities had a few issues.

There is also a possibility that the performance may be affected by gender differences, as reported by Nguyen et al. [49] in their investigation about the moderating role of gender in determining the relationship between personality and academic performance. In their study, they found that emotional stability (the reverse of the neuroticism factor) positively and significantly predicted academic performance of male students, but the same prediction did not occur for female students.

A large and growing body of literature has investigated the effect of personality composition towards team effectiveness [46],[50]. In one of the meta-analytic studies, Bowers et al. [5] investigated whether the teams that were homogeneous in personality outperformed the teams consisting of heterogeneous personality and the findings showed a partial support for heterogeneous teams. They also suggested that effective team personality composition was highly dependent on the type of task, the difficulty level, and the level of communication required in performing the task [5]. Other authors [46],[19] also pointed out that task type played a significant role in determining effective personality composition. While these studies were conducted mostly in the psychology and business domains, further research should be done to investigate the personality composition affecting PP's effectiveness as a pedagogical tool. The issue of whether homogeneity or heterogeneity of personality is good for PP has not been clearly solved yet.

Another possible explanation for our results to support the null hypothesis might be related to other confounding factors such as the skill level of paired students. It would be useful to investigate which factor was the strongest predictor of PP's effectiveness; for instance, by investigating the correlation between personality traits and skill levels towards performance of paired teams. This is because personality factor might not be a strong predictor of PP success as compared with the skill level among paired students. Results from our systematic review revealed that PP worked best when the skill level gap between partners was not too broad [53].

Our results also showed that PP helped students achieve high satisfaction and great confidence in learning programming. Overall, 88% of the paired students indicated that they were happy working with their partner. These results are consistent with those of existing studies that investigated PP's effectiveness [44], [45], [54], [56].

F. Threats to Validity

There are some uncontrolled variables which may have affected the validity of the experimental results. One of these was students' previous programming experience. We noticed that some of the students who already had a few years of programming experience achieved high scores in their test, but scored somewhat low on conscientiousness. They were strong programmers with appropriate knowledge and know-how of programming as compared with other students. Being highly conscientious may not be necessarily for these students in order to obtain good academic results in this particular course.

Another threat was with regard to the change of partners during the tutorials due to a partner's absenteeism. Some students failed to turn up to their allocated tutorial and attended a different tutorial without informing the tutor. This created an unbalanced number of pairs between groups and the likelihood that some students in the controlled group to be moved to the experimental group. The small sample size used in this study (49 students) may also have affected the significance of the results. Cook & Campbell [16] mentioned that when the sample size is large enough, even very small effects can be statistically significant. Since our study only considered conscientiousness as a personality factor, there is a possibility that students' academic performance may also have been affected by other personality variables (e.g.

extraversion, openness to experience, etc.); hence this is one of the topics of our future work.

VI. CONCLUSIONS AND FUTURE WORK

The focus of this study was to determine whether differences on students' personality profile during pairing activities would impact their academic performance. The results of the formal experiment showed a positive correlation between conscientiousness and assignments' scores (r=0.27), and the students' test performance was found to be positively correlated with openness to experience (r=0.35).

The current study did not reject the null hypothesis, thus did not provide any evidence for distinguishing the performance of paired students between personality groups ($p = 0.93$, CI=95%). On average, 88% of students were satisfied with the PP experience. Similarly, most of the students (87%) responded that their confidence level increased when working in pairs. The evidence from this study suggests that regardless of the variation in students' personality disposition, PP not only caused the increase of satisfaction and confidence level, but also brought enjoyment to the class and enhanced students' learning motivation.

In summary, the current findings add to our understanding of the effect of personality variation towards students' academic performance when practicing PP. One of the major implications is to further investigate personality traits of paired students focusing on conscientiousness, openness to experience and extraversion factors. These three factors are also considered educationally relevant [17]. As part of our future work, we will replicate this experiment during the fall 2009. As mentioned by Basili [3], experiments should be replicated externally to strengthen the validity of the earlier studies. Our replication will involve investigating the moderating role of gender, in addition to personality factor.

REFERENCES

[1] M.R. Barrick, M.K. Mount, "The Big Five Personality Dimensions and Job Performance: A Meta-Analysis", Personality Psychology, vol. 44, 1991, pp. 1 – 26.

[2] M.R. Barrick, G.L. Stewart, M.J. Neubert, and M.K. Mount, "Relating Member Ability and Personality to Work-Team Processes and Team Effectiveness'. Journal of Applied Psychology, 1998, 83 (3), pp. 377-391.

[3] V.R. Basili, F. Shull, and F. Lanubile, Building knowledge through families of experiments. IEEE Trans. on Software Engin., 1999, 25 (4), pp. 456-473.

[4] K. Beck Extreme Programming Explained: Embrace Change. Boston: Addison-Wesley, 2000.

[5] C.A. Bowers, J.A. Pharmer and E. Salas, "When Member Homogeneity is Needed in Work Teams: A Meta-Analysis", Small Group Research, 2000, vol. 31 (3), pp. 305-327.

[6] J.H. Bradley and F.J. Hebert, "The Effect of Personality Type on Team Performance", Journal of Management Development, 1997, vol.16 (5), pp. 337-353.

[7] T. Buchanan, J.A. Johnson and L.R. Goldberg, "Implementing a Five-Factor Personality Inventory for Use on the Internet", Journal of Psychological Assessment, 2005, vol. 21 (2), pp. 115-127.

[8] G. Burch and N. Anderson, "Personality as a Predictor of Work-related Behavior and Performance: Recent Advances and Directions for Future Research" in Hodgkinson, G.P. and Ford, J.K. eds. Int'l Review of Industrial and Organizational Psychology, John Wiley & Sons, Ltd, 2008, pp. 261-305.

[9] V.V. Busato, F.J. Prins, J.J. Elshout & C. Hamaker, 'Intellectual ability, Learning Style, Personality, Achievement Motivation and Academic Success of Psychology Students", Higher Education. Personality and Individual Differences, 2000, vol. 29(6), 1057-1068.

[10] T. Chamorro-Premuzic & A. Furnham, "Personality, Intelligence and Approaches to Learning as Predictors of Academic Performance", Personality and Individual Differences, 2008, 44(7), pp. 1596–1603.

[11] J. Chao and G. Atli, "Critical Personality Traits in Successful Pair Programming", AGILE'06, IEEE Comp. Society, 89-93, 2006.

[12] K.S. Choi, F.P. Deek and I. Im, "Exploring the underlying aspects of pair programming: The impact of personality", Information and Software Technology, Oct. 2008, vol. 50(11), pp. 1114-1126.

[13] A. Cockburn, Agile Software Development, Addison-Wesley, Reading, MA., 2001.

[14] A. Cockburn & L. Williams, "The Costs and Benefits of Pair Programming" in Extreme Programming Examined, Succi, G. & Marchesi, M. ,Eds.: Addison-Wesley, 2001.

[15] M.A. Conard, "Aptitude is not enough: How personality and behavior predict academic performance", Journal of Research in Personality, 2006, vol. 40, pp. 339-346.

[16] T.D. Cook and D.T. Campbell, Quasi-experimentation: Design & Analysis issues for field settings. Rand McNally College Publishing, Chicago, 1997.

[17] B. De Raad and H.C. Schouwenburg, "Personality in learning and education", Review. European Journal of Personality, 10. 303-336.

[18] A. Dick and B. Zarnett, "Paired Programming and Personality Traits", XP2002, Italy, 2002.

[19] J.E. Driskell, E. Salas, F.F. Goodwin & P.G. O'Shea, "What Makes a Good Team Player? Personality and Team Effectiveness", Group Dynamics: Theory, Research, and Practice, 2006, 10(4), 249-271.

[20] A. Duff, E. Boyle, K. Dunleavy & J. Ferguson, The Relationship Between Personality, Approach to Learning and Academic Performance. Personality and Individual Differences, 2004, 36(8), pp. 1907–1920.

[21] J. Erickson, K. Lyytinen and K. Siau, "Agile Modeling, Agile Software Development, and Extreme Programming", Journal of Database Management, 2005,16 (4), pp. 88-100.

[22] T. Farsides and R. Woodfield, "Individual Differences and Undergraduate Academic Success: The Roles of Personality, Intelligence, and Application", Personality and Individual Differences, 2003, 34 (7), pp. 1225-1243.

[23] R. Feldt, L. Angelis and M. Samuelsson, "Towards Individualized Software Engineering: Empirical Studies Should Collect Psychometrics", CHASE'08, May 2008, pp. 49–52.

[24] A. Furnham, "The Big Five Vs the Big Four: The Relationship between Myers-Briggs Type Indicator (MBTI) and NEO-PI five factor model of personality", Personality and Individual Differences, 1996, 21 (2), pp. 303-307.

[25] E.F. Gehringer, "A Pair-Programming experiment in a Non-Programming courses", OOPSLA'03, Anaheim, California, USA, 2003, ACM Press, pp.187 - 190.

[26] H. Gevaert, Pair programming unearthed, M.S thesis, University of Manitoba (Canada), Canada, 2007.

[27] L.R. Goldberg, "A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models'. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), Personality Psychology in Europe, Tilburg, The Netherlands: Tilburg University Press, 1999, vol. 7, pp. 7-28.

[28] L.R. Goldberg, J.A. Johnson, H.W. Eber, R. Hogan, M.C. Ashton, C.R. Cloninger & H.C Gough, "The International Personality Item Pool and the future of public-domain personality measures", Journal of Research in Personality, 2006, vol. 40, pp. 84-96.

[29] S.D Gosling, S. Vazire, S. Srivastava and O.P. John, "Should We Trust Web-Based Studies? A Comparative Analysis of Six Preconceptions About Internet Questionnaires", American Psychologist, February 2004, vol. 59 (2), pp. 93-104.

[30] Hanks, B., Wellington, C., Reichlmayr, T. and Coupal, C. 'Integrating Agility in the CS Curriculum: Practices through Values", SIGCSE'08, ACM, Portland, Oregon, USA, 2008.

[31] Hedin, G., Bendix, L. and Magnusson, B. "Teaching Software Development Using Extreme Programming" in Reflections on the Teaching of Programming, 2008, pp.166-189.

[32] S. Heiberg, U. Puus, P. Salumaa and Seeba, "A. Pair-programming effect on developers productivity". XP and Agile Processes in Software Engineering. 4th Int'l Conf., XP 2003. Proc. LNCS Vol.2675. Springer-Verlag, 2003, pp. 215-224.

[33] J. Highsmith, Agile Software Development Ecosystems, Addison-Wesley, 2002.

[34] C.-w, Ho, Examining Impact of Pair Programming on Female Students. Department of Computer Science. Raleigh, NC, North Carolina State University, 2004.

[35] J. Johnson, The IPIP-NEO Personality Assessment Tools. Available online: http://www.personal.psu.edu/j/5/j5j/IPI/ (accessed July 2008)

[36] J.S. Karn and T. Cowling, "A Follow up Study of the Effect of Personality on the Performance of Software Engineering Teams", ACM ISESE'06, 2006.

[37] N. Katira, L. Williams, E. Wiebe, C. Miller, S. Balik. and E. Gehringer, "On understanding compatibility of student pair programmers", ACM. SIGCSE Bulletin, 2004,36 (1), pp. 7-11.

[38] N. Katira, L. Williams and J. Osborne, "Towards Increasing the Compatibility of Student Pair Programmers", ICSE'05 - 27th Int'l Conf. on Software Engin., ACM Press, IEEE Comp. Society, St Louis, Missouri, USA, 2005, 625-626.

[39] S.L. Kichuk and W.H. Wiesner, "The Big Five Personality Factors and Team Performance: Implications for Selecting Successful Product Design Teams", Journal of Engin. and Technology Management, 1997,14, pp. 195-221.

[40] L. Layman, "Changing students' perceptions: an analysis of the supplementary benefits of collaborative software development". Proc. 19th Conf. on Software Engin. Education & Training. IEEE Comp. Society, 2006, pp. 159 – 166.

[41] N.L. Leech, K.C. Barrett and G.A. Morgan, "SPSS for intermediate statistics: use and interpretation" [electronic resource] Mahwah, N.J. Lawrence Erlbaum Associates, 2005.

[42] R.R. McCrae and Jr, P.T.C. "Reinterpreting the Myers-Briggs Type Indicator From the Perspective of the Five-Factor Model of Personality", Journal of Personality, 1989, 57 (1), pp. 17-40.

[43] R.R. McCrae, R.R. and O.P. John, "An Introduction to the Five-Factor model and its application", Journal of Personality, 1992, vol. 60 (2), pp. 175-215.

[44] E. Mendes, L. B. Al Fakhri, A. Luxton-Reilly, "Investigating Pair Programming in a 2nd year Software Development and Design Computer Science Course", Proc. ITiCSE'05, 2005, pp. 296-300.

[45] E. Mendes, L. Al-Fakhri, and A. Luxton-Reilly, "A Replicated Experiment of Pair-programming in a 2nd-year Software Development and Design Computer Science Course," in Proc. 11th Annual SIGCSE Conf. on Innovation and Technology in Comput. Science Educ., ITiCSE06, 2006, pp. 108-112.

[46] S. Mohammed and L.C. Angell, "Personality Heterogeneity in Teams: Which Differences Make a Difference for Team Performance?", Small Group Research, vol. 34 (6), pp. 651-677.

[47] G.A. Morgan, N.L. Leech, G.W. Gloeckner and K.C. Barrett, SPSS for Introductory Statistics. Use and Interpretation. Lawrence Erlbaum Associates, Inc., New Jersey, 2004.

[48] I. Myers-Briggs, M. H. McCaulley, N. L. Quenk, and A. Hammer, MBTI Manual (A Guide to the Development and use of the Myers Briggs Type Indicator), 3rd ed edition ed. vol. Consulting Psychologists Press, 1998.

[49] N.T. Nguyen, L.C. Allen & K. Fraccastoro, "Personality Predicts Academic Performance: Exploring the moderating role of gender", Journal of Higher Education Policy and Management,2005, 27(1), 105 – 116.

[50] M.A.G. Peeters, H.F.J.M.V. Tuijil, C.G. Rutte and I.M.M.J. Reymen, "Personality and Team performance: A Meta-Analysis", European Journal of Personality, 20. 377-396.

[51] A.R. Peslak, "The Impact of Personality on Information Technology Team Projects", SIGMIS-CPR, 2006, pp. 273–279.

[52] V. Pieterse and D.G. Kourie, "Software Engineering Team Diversity and Performance", Proc. of SAICSIT, 2006, pp.180–186.

[53] N. Salleh, "A Systematic Review of Pair Programming Research - Initial Results", Proc. New Zealand Computer Science Research Student Conference (NZCSRSC08), Christchurch, 2008.

[54] P. Sfetsos, I. Stamelos, L. Angelis and I. Deligiannis, "Investigating the impact of personality types on communication and collaboration-viability in pair programming - an empirical study", XP and Agile Processes in Software Engin, XP 2006, Proc. LNCS, Springer-Verlag, 2006, vol.4044, pp. 43-52.

[55] P. Sfetsos, L. Angelis and I. Stamelos, "Investigating the extreme programming system - an empirical study", Empirical Software Engin., 11 (2). 269-301.

[56] L. Williams, and R.R Kessler, "The effects of "pair-pressure" and "pair-learning" on software engineering education", Proc.13th Conf. on Software Engin. Education & Training, 2000, pp. 59–65.

[57] L. Williams, L. Layman, J. Osborne and N. Katira, "Examining the Compatibility of Student Pair Programmers", AGILE 2006 Conf. (AGILE'06), IEEE Comp. Society, 2006.