

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

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

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

John Grundy  
Center for Complex  
Software Services  
Swinburne University of  
Technology, Australia  
jgrundy@swin.edu.au

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

## ABSTRACT

This paper describes a formal experiment carried out to investigate the effect of the Personality factor conscientiousness on the effectiveness of Pair Programming as a pedagogical tool in higher Education. This experiment took place at the University of Auckland, using as subjects undergraduate students attending an introductory programming course. Conscientiousness was chosen because it has been shown to be the most consistent predictor of academic achievement. Our findings showed that differences in conscientiousness level did not significantly affect the academic performance of students who pair programmed, which could have been due to the short duration of the tasks used throughout the experiment. However, results revealed that another Personality factor - Openness to experience - presented a significant correlation with paired students' academic performance.

## Categories and Subject Descriptors

k.3.2 [Computer and Information Science Education]:  
Computer Science Education

## General Terms

Management, Human Factors.

## Keywords

Pair programming, formal experiment, personality type, five-factor model, conscientiousness, empirical investigation, higher education.

## 1. INTRODUCTION

Pair programming (PP) has been recognized as one of the key practices in the Extreme Programming development methodology, and becoming more prevalent in industry as well as in educational settings [4], [27]. It involves two people working together side-by-side on the same set of design or coding tasks; throughout the process they alternate their role as the "driver" and "navigator" [56]. Proponents of PP in higher education institutions have investigated its usefulness and effectiveness as a Computer Science/Software Engineering (CS/SE) pedagogical tool (e.g. [39], [40], [44]), and some of these studies reported very

promising results. For example, evidence suggests that PP can improve course completion rate [39], [44] increase students confidence level [39], [28], improve performance on exams [5], [6], and facilitate students to work more efficiently on programming tasks [13], [55].

Based on evidence from our systematic literature review (SR) of PP in higher education [50], we found that only 23% of the included studies have empirically investigated factors that may affect PP's success, including pair compatibility. One of the most significant findings from our SR suggests that personality was one of the most common factors investigated in previous PP studies. Nonetheless, the results from these studies were inconsistent in terms of the effect or influence of personality towards PP's effectiveness [51], [50].

Most of the studies in our SR reported that personality had no significant influence in determining PP's effectiveness [11], [30],[35],[26]. However, since existing literature in Agile methods suggests that developers' personality is one of PP's most critical success factors [14], [31] it seems important in our view to better understand the role and impact of personality in determining PP's effectiveness. The selection of personality traits as variables would provide an advantage in overcoming the problem of bad pairing experiences among students, which were reported in some PP studies [37], [32]. The discomfort or incompatibility experienced working with a partner might be due to a mismatch in psychological aspects such as personality and gender combinations.

This paper reports on a formal experiment conducted with undergraduate students at The University of Auckland during the semester 1, 2009. The aim of our study was to extend our previous study [51] regarding the effects of personality composition on the effectiveness of PP, focusing this time on the conscientiousness factor of the Five-Factor personality model. Conscientiousness is one of the broad personality traits of FFM, which is most closely linked to determination, or will to achieve, persistency, efficiency, and being organized [49]. Since most studies that examined the influence of personality on educational success pointed to conscientiousness as the strongest predictor of performance [15], [8], [58], we hypothesise that this factor is a strong determinant of PP's effectiveness as a pedagogical tool in higher education.

The objective of our research is to improve the effectiveness of PP as a pedagogical tool for CS/SE education by investigating the effects that conscientiousness of paired students may have on PP's effectiveness. The contribution of our study is twofold: i) it contributes to the PP body of knowledge by providing evidence in understanding the role of personality as a determinant of PP's

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ICSE'10, May, 2010, Cape Town, South Africa.

Copyright 2010 ACM 1-58113-000-0/00/0004...\$5.00.

effectiveness, from the perspective of Five-Factor personality model; and ii) it provides evidence that can be used to ameliorate CS/SE learning in higher education institutions.

In the next Section (Section 2), we summarize the related work involving investigation of personality aspects in PP and related literature in psychology regarding team personality composition. In Section 3, we briefly describe the Five-Factor Model as a personality framework used in this study. Section 4 presents the research methodology followed by the results in Section 5. Section 6 presents the discussion of our findings and finally we draw our conclusions in Section 6.

## 2. RELATED WORK

Personality has been identified as a potentially important variable in ensuring successful implementation of PP [52]. This is due to the nature of PP which involves collaboration or interaction

between two programmers working together on a single machine. Thus the practice's success is largely determined by how effectively they work as a team, despite their skills and abilities. In this regard, considerable number of studies have been conducted to understand the role of personality in determining PP's effectiveness either in academic or industry context [53], [29], [51].

The summary of studies relating personality to PP can be seen in Table 1. In general, findings from these studies were quite diverse and thus inconclusive whether personality could significantly affect the outcome or productivity of pair programmers. Only two studies [53], [12] presented positive findings, reporting that paired students of different personality types performed better when compared with paired students of similar personality type. Most studies reported that personality had no significant influence in determining PP's effectiveness [11], [30], [35], [26].

**Table 1: List of PP studies investigating Personality factor**

Author(s)	Type of study	Sub.	Size	IV	DV	Outcomes	Personality measurement
Chao & Atfi [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. [30]	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. [34]	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. [35]	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 [37]	Survey	Stud.	119	All paired	Perception towards collaboration	Personality had no significant effect towards perception to collaborate.	MBTI
Sfetsos et al. [53]	Formal Exp.	Stud.	70	Personality	PP's effectiveness	Paired of mixed personalities performed better than paired of same personality.	KTS
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 [17]	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
Hannay et al. [29]	Regression	Prof.	196	Personality	Pair performance	The effects of personality were not consistent and suggest that personality as only a moderate predictor for pair performance.	Big Five
Salleh et al. [51]	Formal Exp.	Stud.	54	Personality	PP's effectiveness	Differences in personality had no significant effect on paired students performance.	Five-Factor Model
Exp – Experiment Sub. – Subject Stud – Student Unk. – Unknown N/A – Not available Prof. – Professional							

Although various personality models were applied in previous PP studies, most studies used the Myers-Briggs Type Indicator (MBTI) as a personality assessment method [34], [35], [37],[57], [12]. MBTI is one of the most popular instruments used to measure an individual's personality types and commonly used in the area of consultancy, and business training [25]. More recently, however, some PP studies have investigated personality using the Five Factor of Big-Five personality models [29], [51]. There is evidence that the Five-Factor personality dimensions are a robust taxonomy of personality and currently considered the predominant taxonomy of personality by personality psychologists [2], [7]. This evidence motivated us to employ the Five-Factor Model (FFM) in our study.

The present study is an extension of our previous work in [51], where PP's team personality composition was investigated. In particular, we investigated how differences in paired students' personality profile affected their academic performance, focusing on the conscientiousness factor part of the FFM. Conscientiousness was chosen, and also used herein, because this factor was reported to be associated with team performance as well as academic success [9], [24], [20], [8]. The results from [51] did not provide strong support to distinguish performance between paired students of similar and mixed personality. Despite a considerably small sample size, we realized that one of the important issues found in [51] related to the pair formation strategy: paired students of mixed personality consisted of students of high and low conscientiousness and such a matching could possibly produce an incompatible pair due to dissimilarities in character or attitude. Thus, in the present study we used a different approach to pair formation in order to overcome this issue.

Research in the Psychology literature has proposed some strategies for operationalizing team composition based on the FFM. The four common methods are based on the aggregation of the mean, the variance, the maximum and the minimum scores of individual personality traits in a team [2]. It has also been reported that the appropriateness of methods for operationalizing team composition very much depends on the nature of the task, the specific traits being analyzed and the research questions being investigated [2]. As the PP practice typically relates to problem solving-based tasks, the mean method seems to be a better method of operationalization [47]. This means that, in terms of the conscientiousness factor, pairs where both partners present a high score of conscientiousness would result in a better performance as both partners possess positive attributes such as being a high achiever, organized, responsible, and diligent.

In terms of academic success, existing literature implies a consistently positive relationship between conscientiousness and students' academic performance [15], [8], [20], [24], [9], [58]. A recent meta-analysis of the relationship between FFM and academic performance supported the hypothesis that conscientiousness was the strongest predictor of academic performance [48]. It has been reported in [24] that volitional facets of conscientiousness such as self-discipline, achievement striving, and dutifulness are important characteristics for academic achievement. With regard to team performance, results

from a meta-analysis of personality and team performance suggested that teams should be composed of members that are highly and similarly conscientious [47]. Therefore, we were interested to investigate the effects of conscientiousness when applied to PP focusing on an academic or student learning context in tertiary institutions. In the following Section, we introduce the FFM theoretical framework, which is the basis of the personality assessment used in our study.

### 3. FIVE-FACTOR MODEL

The Five-Factor Model (FFM), as its names implies, comprises five broad traits - Openness to experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism – and provides a structure that categorizes dimensions of differences in human personality [38]. Openness to experience describes intellectual, cultural, or creative interest. Conscientiousness is concerned with one's achievement orientation. Persons who are high on conscientiousness tend to be hardworking, organized, able to complete tasks thoroughly, and 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 who is a low neurotic tends to appear calm, confident, and secure, whereas a highly neurotic individual tends to be moody, anxious, nervous, and insecure [19].

The selection of FFM as personality assessment framework in our study was due to its comprehensive nature and its ability to capture the basic temperament and dispositional factors relevant to the educational context [49]. In terms of its validity and reliability, FFM has been generally accepted by personality psychologists who suggest that such a broad trait of dimensions adequately represents human personality attributes [1],[2].

### 4. RESEARCH METHODOLOGY

This Section details the formal experiment conducted during semester 1, 2009 at the University of Auckland.

#### 4.1 Research Objectives

The 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 as follows:

*Object of study:* PP technique.

*Purpose:* To improve the effectiveness of PP as a pedagogical tool in higher education institutions in order to increase students' amount of learning and satisfaction.

*Focus:* To investigate the influence of conscientiousness factor of FFM personality model 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.

**Table 2: GQM Definition**

Goal(s)	Question(s)	Metric(s)
To investigate the effect of conscientiousness towards a successful pair configuration	Do differences in conscientiousness level within a pair affect the pair's academic performance?	Students' academic performance measured by assignments, test and final exam 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

## 4.2 Context

The formal experiment was conducted during semester 1, 2009 involving first year undergraduate students enrolled in an introductory programming course. Students paired during compulsory weekly tutorials, run by a tutor and a few TAs.

## 4.3 Hypotheses

According to the FFM, the level of conscientiousness a person has indicates the degree of aspiration or one's desire for achievement [38]. Therefore, a highly conscientious individual tends to be more persistent, responsible, organized, thorough, and ambitious. In contrast, persons low on conscientiousness are expected to be impulsive, irresponsible, disordered, and to lack a desire for achievement [18].

Research on personality suggests that students' academic performance is positively associated with their level of conscientiousness [8], [58], [48]. It has also been reported in the literature that teams comprising a higher average score of conscientiousness demonstrated better job performance [2], [45]. Extending this logic into the realm of PP, we predicted that pairs consisting of highly conscientious students are expected to achieve better academic performance than pairs presenting low levels of conscientiousness. Hence, the following null hypothesis investigated in our study is as follows:

*H<sub>0</sub>: Differences in conscientiousness level do not affect the academic performance of students who pair programmed.*

which is contrasted by the following alternative hypothesis:

*H<sub>A</sub>: Differences in conscientiousness level affect the academic performance of students who pair programmed.*

Table 3 shows the categorization of pairs according to students' level of conscientiousness. Pair (C<sub>High</sub>, C<sub>High</sub>) denotes a combination where both students have high levels of conscientiousness. Our experiment also looked into the association between an individual student's personality score with their academic performance, and level of satisfaction and confidence when working in pairs.

**Table 3: Pair Configuration**

Experimental Group	Controlled Group
Pair (C <sub>High</sub> , C <sub>High</sub> )	Pair (C <sub>Low</sub> , C <sub>Low</sub> )

Pair (C <sub>High</sub> , C <sub>High</sub> )	Pair (C <sub>Med</sub> , C <sub>Med</sub> )
Pair (C <sub>Med</sub> , C <sub>Med</sub> )	Pair (C <sub>Low</sub> , C <sub>Low</sub> )

## 4.4 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 selected to measure PP's effectiveness were "academic performance" and students' "satisfaction". Hence, PP's effectiveness and satisfaction were our dependent variables and level of conscientiousness our independent variable (single-factor).

In this study, PP's effectiveness was measured using assignments, test and exam scores, and satisfaction was measured using a questionnaire where all questions employed a five-point likert-scale.

## 4.5 Instrumentation and Materials

During the first course lecture, one of the authors provided 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 (The test can be accessed at this public domain URL: <http://www.personal.psu.edu/j5j/IPIP/>). 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 [33]. The IPIP-NEO consists of 120 items which descriptions were authored by John A. Johnson [33]. The test produces scores in a numerical scale, where 0 and 99 represent the lowest and the highest scores for each factor, respectively. Based on the distribution of scores for the conscientiousness trait, the grouping was done to provide a more balanced number of subjects within each group. As such, the classification of conscientiousness level used in our study is shown in Table 4.

**Table 4: Personality Scores Level**

Scores	Lowest 40%	Middle 30%	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. In line 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 students enrolled in an introductory programming course (COMPSCI101). During the course, students learnt about basic programming concepts 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

## 4.6 Experimental Procedure

We followed the same procedure carried out in our previous study [51], where each of the tutorial sessions was treated as an independent formal experiment. During the first week of the semester, students' personality data were gathered using the online IPIP test. The results of the personality test were then used to allocate partners. For this purpose, the personality scores relating to the conscientiousness factor were used to assign students to one of three possible groups: low, medium or higher level of conscientiousness.

The allocation of pairs within each group was done randomly. Since we only had conscientiousness as our sole independent variable, our hypothesis was investigated using a "single factor between-group design" [43] experimental design.

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 a mid-term test were also graded but completed individually.

The outcomes measured from the experiment were the students' academic performance in their test, final exam, 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.

## 4.7 Analysis Procedure

In order to test our null hypothesis, we used a one-way analysis of variance (ANOVA) to analyse the difference in academic performance between groups. We used a Kruskal-Wallis test 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 employed was SPSS v. 17.

# 5. RESULTS

## 5.1 Subjects

A total of 453 students enrolled in the COMPSCI 101 course in semester 1, 2009. Of these, 295 (65.1%) planned to obtain a BSc, 44 (9.7%) a BArts, 22 (4.9%) a BCom, 6 (1.5%) a Graduate Diploma of Science degree, and the remaining to obtain a Bachelor of Law co-joint with Commerce and Science.

There were 350 male students (74%), and subjects' age ranged from 19 to 52 years (median age = 19 years). Of the 317

students who responded to the demographic survey, 85% indicated that they did not have any work experience. Of the 453 students, 218 students (48%) completed the personality test. Therefore, the sample size used in our analysis was 218.

## 5.2 Correlation between conscientiousness and Academic Performance

Figure 1 shows the distribution of assignments' scores according to students' level of conscientiousness. The scores between groups were somewhat similar, although on average, paired students of high conscientiousness obtained a slightly higher mark than the other groups. Note that each group had a few outliers, representing students who did not submit some of the assignments.

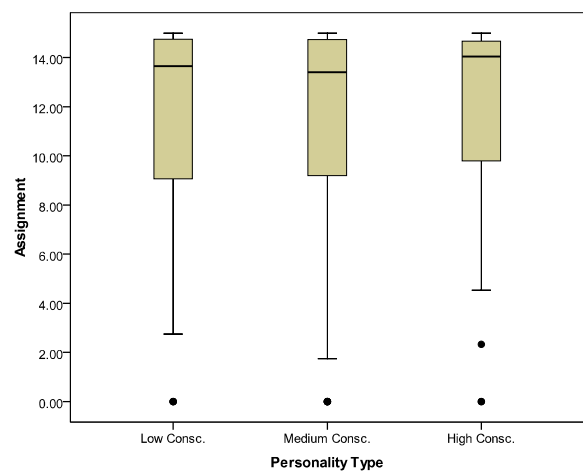
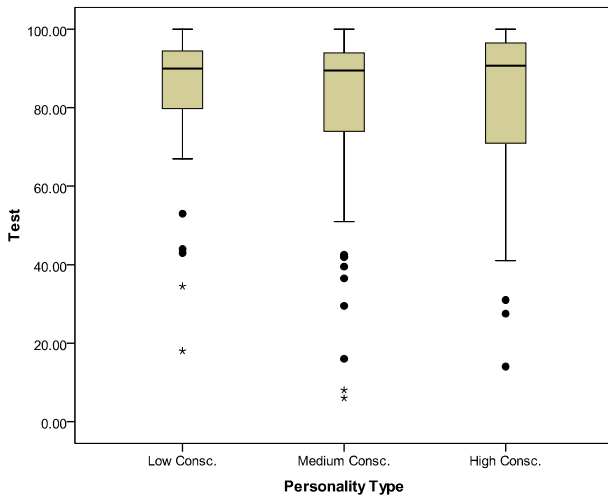


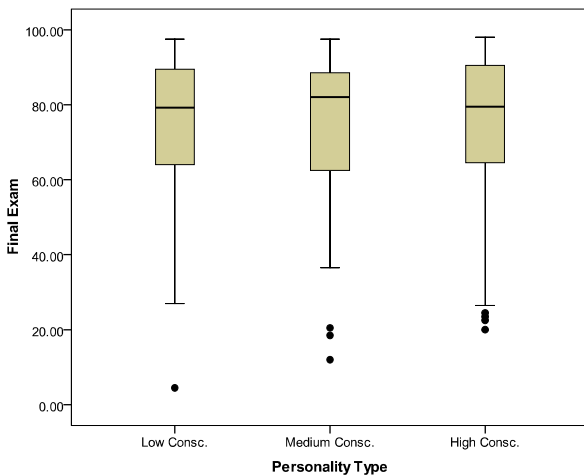
Figure 1: Comparison of assignments scores between groups

In terms of students' test scores, the dispersion of scores was substantially larger for students in the high conscientiousness group, suggesting this group was more heterogeneous compared to the other two groups (see Figure 2). The median scores for each group were similar, thus suggesting that level of conscientiousness may not necessarily determine/be related to test performance, at least for the sample employed.



**Figure 2 Comparison of test scores between groups**

Figure 3 shows the boxplots of final exam scores for each group. The Low conscientiousness group presented the flattest distribution of scores; however the median scores and the upper quartiles between groups were also similar, thus suggesting, at least based on the boxplots, that there were no noticeable differences in final exam scores between groups



**Figure 3: Comparison of final exam scores between groups**

Table 5 shows the results from applying bivariate correlation test to measure the association between FFM variables and academic performance. Contrary to our expectations, there are no significant relationships between level of conscientiousness and performance. However, students' performance (assignments, test, and final exam scores) showed a significant positive relationship with Openness to experience. The strongest positive correlation was between final exam scores and openness to experience,  $r(182) = 0.22, p = 0.003$ . The findings regarding Openness to experience were consistent with those from our previous study [51].

**Table 5: Correlation between the academic performance and the big-five traits (N=184)**

	1	2	3	E	A	C	N	O
1	1	0.43**	0.60**	0.02	-0.00	-0.03	0.02	<b>0.21**</b>
2		1	0.83**	-0.07	0.04	-0.11	-0.04	<b>0.13*</b>
3			1	-0.07	0.03	-0.08	-0.04	<b>0.22**</b>
E				1	0.02	0.35**	-0.24**	0.11
A					1	0.33**	0.06	0.21**
C						1	-0.14*	0.11
N							1	-0.04
O								1

1. Assignments 2. Test 3. Final Exam (E) Extraversion (A) Agreeableness (C) Conscientiousness (N) Neuroticism (O) Openness to Experience

\*\* . Correlation is significant at the 0.01 level (2-tailed).

\* . Correlation is significant at the 0.05 level (2-tailed).

### 5.3 Hypothesis Testing

We tested the hypotheses using One-way analysis of variance (ANOVA) to analyze whether there was any significant difference in academic performance between the three levels of conscientiousness (low, medium, and high). One-Way ANOVA procedure was reported to be robust and can be relied upon even when distributional assumptions are violated [43].

Table 6 provides the mean values and standard deviation values for academic performance for each group. Mean differences are almost similar for all measures of PP's effectiveness. The Levene test showed that the variances between groups were not significant; therefore the assumption for variance homogeneity was not violated. The overall  $F$  values for the three ANOVA are presented in Table 7. The results showed that there was no significant difference in academic performance between the three groups of conscientiousness (i.e.  $F(2, 215) = 0.56, p = 0.57$ , for assignments;  $F(2, 212) = 0.31, p = 0.74$ , for test;  $F(2, 206) = 0.06, p = 0.94$ , for final exam). Since none of the  $F$  values were statistically significant, no post-hoc analysis was needed. Our results indicated that we could not find strong support to reject the null hypothesis. Thus, based on our data, we found that PP's effectiveness was not affected by differences in conscientiousness levels among paired students.

**Table 6: Mean and standard deviation of paired students**

	Conscientiousness Level	N	Mean	SD
Assignments	Low Consc	53	11.26	4.50
	Medium Consc.	94	11.32	4.34
	High Consc.	71	11.94	3.88
	Total	218	11.51	4.23
Test Scores	Low Consc	52	83.54	17.41
	Medium Consc.	93	80.83	21.21
	High Consc.	70	82.05	20.43
	Total	215	81.88	20.04
Final Exam	Low Consc	50	74.21	20.67
	Medium Consc.	90	74.40	19.44
	High Consc.	69	73.32	21.53
	Total	209	73.99	20.35

**Table 7: ANOVA Results**

		Sum of Squares	Df	Mean Squares	F	Sig.
Assign.	Between Groups	20.00	2	10.00	0.56	0.57
	Within Groups	3865.01	215	17.97		
	Total	3885.01	217			
Test	Between Groups	246.93	2	123.46	0.31	0.74
	Within Groups	85662.67	212	404.07		
	Total	85909.59	214			
Final Exam	Between Groups	48.75	2	24.38	0.06	0.94
	Within Groups	86084.90	206	417.89		
	Total	86133.66	208			

## 5.4 Results on Satisfaction and Confidence

In order to measure paired students' levels of satisfaction and confidence, we distributed PP surveys in each tutorial session. Altogether there were nine weeks of tutorials starting from the second week of the semester until the final week of tutorials. Data were analyzed separately as each tutorial was treated as a single independent "mini-experiment".

**Table 8: Mean Rank for Satisfaction Level**

	Consc. Level	N	Mean Rank	Sig.	Satisfied/Very Satisfied (%)
Tut. 2 N=146	Low	38	57.68	0.00	89
	Medium	52	68.73		
	High	56	88.61		
Tut. 3 N=131	Low	27	58.46	0.34	90
	Medium	47	65.76		
	High	57	70.62		
Tut. 4 N=156	Low	32	68.02	0.06	87
	Medium	68	75.35		
	High	56	88.32		
Tut. 5 N=132	Low	36	57.25	0.16	92
	Medium	43	69.12		
	High	53	70.66		
Tut. 6 N=132	Low	26	67.00	0.28	94
	Medium	36	59.85		
	High	67	70.18		
Tut. 7 N=115	Low	28	59.96	0.02	87
	Medium	42	48.19		
	High	45	65.93		
Tut. 8 N=106	Low	28	51.52	0.49	93
	Medium	35	50.73		
	High	43	57.05		
Tut. 9 N=94	Low	18	46.11	0.44	87
	Medium	34	44.01		
	High	42	50.92		
Tut.10 N=91	Low	16	49.75	0.76	93
	Medium	37	45.55		
	High	91	44.86		

The response rate of the post-experimental survey was of approximately 67% during the first week, but decreased to 42% for the last week of tutorials. Our data analysis showed that on average 111 (90.2%) out of an average of 123 students attending the tutorials (see Table 8) were satisfied working with their partner (more than 50% reported that they were very satisfied); in addition, approximately an average of 107 students who attended the tutorial classes (see Table 8) answered that their level of confidence solving tasks with their partner was high by working in pairs.

The Kruskal-Wallis Test was used to compare satisfaction levels between conscientiousness groups. This statistical test was chosen because our dependent variable was measured on an ordinal scale. Table 8 shows the mean satisfaction rank of paired students. The group with the highest mean rank had the highest level of satisfaction. Although paired students of high conscientiousness appeared to score higher ranks in most of the experiment units, these differences were not always significant. As can be seen in Table 8, only Tutorial 2 and 7 showed a significant value ( $p = 0.00$  and  $p=0.02$  respectively) but overall results demonstrated that the satisfaction level of paired students were not affected by students' level of conscientiousness.

## 6. DISCUSSION

The focus of our study was to investigate the effect that level of conscientiousness may have on PP's effectiveness as a pedagogical tool. Our results showed that the academic performance of paired students was not significantly affected by their level of conscientiousness. Although these results seem to contradict some of the previous findings reported in the literature [58], [8], [9], there is also evidence showing that level of conscientiousness may not be always prominent in affecting the performance of teams of students [42], [36], [47].

Peeters et al. [47] argued that the effect of conscientiousness can be absent in student teams due to the short period of time available for teams to complete a task and also due to the low levels of interdependency among team members. As our students practiced PP only once a week during tutorials that lasted for less than two hours, this might explain why we did not find a significant difference between the three different of conscientious levels and performance.

An investigation into self-managed groups also did not find a significant relationship between conscientiousness and team outcome [59]. They suggested that "*conscientiousness may become less important in team-based tasks because groups are able to recognize and compensate for the lack of conscientious individuals*" (pg. 76) [59]. In another study, Le Pine et al. [60] showed that there were differential effects for facets of conscientiousness in teams' decision making performance. In their study, the effects of conscientiousness were found opposite than what they have expected, and their further analysis showed that the findings were due to the traits reflecting a dependability facet rather than the achievement facet. Thus, we may also need to carry out a supplementary analysis to find out the reasons behind the results we obtained in our study.

Another possible explanation for our findings relates to the fact that the assessment of tasks/assignments contributed towards a student's course grade, therefore students may have tended to

perform well regardless of their personality attributes. As reported by Kichuk, “*The relative novelty of being a university student and the perceived consequences of not performing well may have caused most students to behave conscientiously while doing the task regardless of how they scored on the personality profile*” (pg. 211) [36]. This can also explain why team composition based on personality traits differed between academic (i.e. lab setting) and industry environments (i.e. actual field setting) in relation to team performance, as reported in [5].

Of all the FFM personality trait measures, our results showed that ‘Openness to experience’ showed the most prominent relationship with PP’s effectiveness (measured by academic performance). These findings were consistent with those found in our previous study [51], and also corroborate results from other studies [22], [10], [24].

Previous research [6], [54] suggests that ‘Openness to experience’ facilitates the use of learning strategies, and those students with a relatively high level of ‘Openness to experience’ were described as being foresighted, intelligent, and resourceful [49]. It was also expected that ‘Openness to experience’ to be more influential on performance for the tasks that require creativity or tackling of abstract problems [18]. Within the context of our study, paired students were given programming exercises typically considered as abstract by their nature. Therefore, this may explain why in our case ‘Openness to experience’ may have influenced PP’s effectiveness far more than the conscientiousness aspect. As part of our future work we will investigate the effects of ‘Openness to experience’ towards PP’s effectiveness.

### 6.1 Threats to the Validity of Our Results

There are several potential threats to the validity of our results. The first relates to the sample size – 218 students. A larger sample would have helped statistical power of the results [41], [21].

Another limitation relates to the construct validity of our dependent variable. Herein we used a student’s academic performance as a surrogate measure of PP’s effectiveness. However students’ performance may also be affected by their cognitive or mental ability. Regardless of one’s personality behavior while pairing, students may perform well due to their ability or competency in programming. However, since the aim of this study is to increase the amount of students’ learning due to practicing PP throughout the entire semester, measuring their academic performance is in our view appropriate to be used in our context.

A further limitation is that our study did not control for the effects of gender. Although earlier meta-analysis suggested that gender may affect personality traits [23], secondary analyses of personality data based on FFM reported that the difference was generally subtle or small relative to individual personality variation within a single gender group [16]. In particular, gender differences only appeared pervasive in facets of neuroticism and agreeableness, and fairly negligible for conscientiousness [16]. Therefore, this issue might not play a significant role in the results obtained in our study.

## 7. CONCLUSIONS AND FUTURE WORK

In summary, the current study did not reject the null hypotheses, thus did not provide any evidence for distinguishing the performance of paired students between different levels of conscientiousness. Despite the counterintuitive findings regarding the effects of conscientiousness, the results of the formal experiment showed a positive correlation between Openness to experience and all measures of PP’s effectiveness thus corroborating findings of existing studies in [10], [45], [22]. On average, 88% of students were satisfied with the PP experience. Similarly, 87% of students responded that their confidence level increased when working in pairs. This evidence suggests that regardless of the variation in students’ conscientiousness level, PP not only caused the increase of satisfaction and confidence level, but also brought enjoyment to the class and enhanced students’ learning motivation.

The current findings add to our understanding of the effect of conscientiousness towards students’ academic performance when practicing PP. As part of our future work, we will further investigate personality traits of paired students focusing on Openness to experience factor as this factor seems to bear the strongest relationship with students’ academic performance.

## 8. ACKNOWLEDGMENTS

The study was funded by the Ministry of Higher Education Malaysia. The author would like to thank Ann Cameron, Adriana, and Jacob for allowing us to run the experiment in their labs. Thanks also to CS101 demonstrators for their assistance during the tutorials, and to all students who participated in the experiment.

## 9. REFERENCES

- [1] M. R. Barrick, and M. K. Mount, “The Big Five Personality Dimensions and Job Performance: A Meta-Analysis,” *Personality Psychology*, vol. 44, pp. 1-26, 1991.
- [2] M. R. Barrick, G. L. Stewart, M. J. Neubert *et al.*, “Relating Member Ability and Personality to Work-Team Processes and Team Effectiveness,” *Journal of Applied Psychology*, vol. 83, no. 3, pp. 377 - 391, 1998.
- [3] V. R. Basili, F. Shull, and F. Lanubile, “Building knowledge through families of experiments,” *IEEE Transaction on Software Engineering*, vol. 25, no. 4, pp. 456-473, Jul-Aug 1999, 1999.
- [4] K. Beck, *Extreme Programming Explained: Embrace Change*: Addison-Wesley, 2000.
- [5] S. T. Bell, “Deep-Level Composition Variables as Predictors of Team Performance: A Meta-Analysis,” *J. of Applied Psychology*, vol. 92, no. 3, pp. 595-615, 2007.
- [6] G. Blickle, “Personality traits, learning strategies, and performance,” *European Journal of Personality*, vol. 10, pp. 337-352, 1996.
- [7] G. S. J. Burch, and N. Anderson, “Personality as a Predictor of Work-related Behavior and Performance: Recent Advances and Directions for Future Research,” *International Review of Industrial and Organizational Psychology*, G. P. Hodgkinson and J. K. Ford, eds., pp. 261-305: John Wiley & Sons, Ltd, 2008.



- [8] V. V. Busato, F. J. Prins, J. J. Elshout *et al.*, "Intellectual ability, Learning Style, Personality, Achievement Motivation and Academic Success of Psychology Students in Higher Education," *Personality and Individual Differences*, vol. 29, no. 6, pp. 1057-1068, 2000.
- [9] T. Chamorro-Premuzic, and A. Furnham, "Personality Traits and Academic Examination Performance," *European Journal of Personality*, vol. 17, pp. 237-250, 2003.
- [10] T. Chamorro-Premuzic, and A. Furnham, "Personality, Intelligence and Approaches to Learning as Predictors of Academic Performance," *Personality and Individual Differences*, vol. 44, no. 7, pp. 1596-1603, 2008.
- [11] J. Chao, and G. Atli, "Critical Personality Traits in Successful Pair Programming," in *Proc. AGILE 2006*. IEEE Computer Society, pp. 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*, vol. 50, no. 11, pp. 1114-1126, 2008.
- [13] D. C. Cliburn, "Experiences with Pair programming at a Small College," *Journal of Computing Sciences in Colleges*, vol. 19, no. 1, pp. 20-29, 2003.
- [14] A. Cockburn, *Agile Software Development*, Boston, MA.: Addison-Wesley Longman Publishing Co. Inc., 2002.
- [15] M. A. Conard, "Aptitude is not enough: How personality and behavior predict academic performance," *Journal of Research in Personality*, vol. 40, pp. 339-346, 2006.
- [16] P. T. Costa, A. Terracciano, and R. R. McCrae, "Gender differences in personality traits across cultures: Robust and surprising findings," *J. of Personality and Social Psychology*, vol. 81, no. 2, pp. 322-331, 2001.
- [17] A. Dick, and B. Zarnett, "Paired programming and personality traits," in *XP2002*, Italy, 2002.
- [18] J. E. Driskell, R. Hogan, and E. Salas, "Personality and group performance," *Rev. Personality Social Psychology*, vol. 9, pp. 91-113, 1987.
- [19] J. E. Driskell, E. Salas, F. F. Goodwin *et al.*, "What Makes a Good Team Player? Personality and Team Effectiveness," *Group Dynamics: Theory, Research, and Practice*, vol. 10, no. 4, pp. 249-271, 2006.
- [20] A. Duff, E. Boyle, K. Dunleavy & J. Ferguson, "The Relationship Between Personality, Approach to Learning and Academic Performance," *Personality and Individual Differences*, 36(8), 1907-1920, 2004.
- [21] T. Dyba, V. B. Kampenes, and D. I. L. Sjoberg, "A Systematic Review of statistical power in Software Engineering Experiments," *Information and Software Technology*, vol. 48, pp. 745-755, 2006.
- [22] T. Farsides, and R. Woodfield, "Individual Differences and Undergraduate Academic Success: The Roles of Personality, Intelligence, and Application," *Personality and Individual Differences*, vol. 34, no. 7, pp. 1225-1243, 2003.
- [23] A. Feingold, "Gender differences in Personality: A Meta-Analysis," *Psychological Bulletin*, vol. 116, no. 3, pp. 429-456, 1994.
- [24] F. D. Fruyt, and I. Mervielde, "Personality and Interests as Predictors of Educational Streaming and Achievement," *European Journal of Personality*, vol. 10, pp. 405-425, 1996.
- [25] 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*, vol. 21, no. 2, pp. 303-307, 1996.
- [26] H. Gevaert, *Pair programming unearthed*, M.S thesis, University of Manitoba (Canada), Canada, 2007.
- [27] B. Hanks, C. Wellington, T. Reichlmayr *et al.*, "Integrating Agility in the CS Curriculum: Practices through Values," in *SIGCSE'08*, Portland, Oregon, USA, 2008..
- [28] B. Hanks, C. McDowell, D. Draper *et al.*, "Program quality with pair programming in CS1," *ACM. SIGCSE Bulletin*, vol. 36, no. 3, pp. 176-80, 2004.
- [29] J. E. Hannay, and E. Arisholm, "Effects of Personality on Pair Programming," *IEEE Transaction on Software Engineering*, (accepted for publication), 2009.
- [30] S. Heiberg, U. Puus, P. Salumaa *et al.*, "Pair-programming effect on developers productivity," *Extreme Programming and Agile Processes in Software Engineering. 4th International Conference, XP 2003. Proceedings (Lecture Notes in Computer Science Vol.2675)*. Springer-Verlag, 2003, pp. 215-24, 2003.
- [31] J. Highsmith, *Agile Software Development Ecosystems*, Boston, MA.: Addison-Wesley Longman Publishing Co. Inc., 2002.
- [32] C.-w. Ho, *Examining Impact of Pair Programming on Female Students*, TR-2004-20, North Carolina State University, Raleigh, NC, 2004.
- [33] J. Johnson. "The IPIP-NEO Personality Assessment Tools," accessed July 2008; <http://www.personal.psu.edu/j5j/IPIP/>.
- [34] N. Katira, L. Williams, E. Wiebe *et al.*, "On understanding compatibility of student pair programmers," *ACM. SIGCSE Bulletin*, vol. 36, no. 1, pp. 7-11, 2004.
- [35] N. Katira, L. Williams, and J. Osborne, "Towards Increasing the Compatibility of Student Pair Programmers," in *ICSE'05 - 27th International Conference on Software Engineering*, St Louis, Missouri, USA, 2005, pp. 625-626.
- [36] S. L. Kichuk, and W. H. Wiesner, "The Big Five Personality Factors and Team Performance: Implications for Selecting Successful Product Design Teams," *Journal of Engineering and Technology Management*, vol. 14, pp. 195-221, 1997.
- [37] L. Layman, "Changing students' perceptions: an analysis of the supplementary benefits of collaborative software development," *Proceedings. 19th Conference on Software Engineering Education & Training. IEEE Computer Society. 2006*, pp. 159-166, 2006.

- [38] R. R. McCrae, and O. P. John, "An Introduction to the Five-Factor model and its application," *Journal of Personality*, vol. 60, no. 2, pp. 175-215, June 1992, 1992.
- [39] C. McDowell, L. Werner, H. E. Bullock *et al.*, "The Impact of Pair Programming on Student Performance, Perception and Persistence." in *Proc. 25<sup>th</sup> International Conf. Software Engineering (ICSE'03)*, pp. 602-607, 2003.
- [40] E. Mendes, L. B. Al-Fakhri, and A. Luxton-Reilly, "Investigating pair-programming in a 2nd-year software development and design computer science course." in *Proc. ITiCSE'05*, Portugal, pp. 296-300, 2005.
- [41] J. Miller, J. Daly, M. Wood *et al.*, "Statistical power and its subcomponents - missing and misunderstood concepts in empirical software engineering research," *Information and Software Technology*, vol. 39, pp. 285-295, 1997.
- [42] S. Mohammed, J. E. Mathieu, and A. L. B. Bartlett, "Technical-Administrative Task Performance, Leadership Task Performance, and Contextual Performance: Considering the influence of Team-and Task-Related Composition Variables," *J. of Organizational Behavior*, vol. 23, pp. 795-814, 2002.
- [43] G. A. Morgan, N. L. Leech, G. W. Gloeckner *et al.*, *SPSS for Introductory Statistics. Use and Interpretation*, Second ed., New Jersey: Lawrence Erlbaum Associates, Inc., 2004.
- [44] N. Nagappan, L. Williams, M. Ferzli *et al.*, "Improving the CS1 experience with pair programming," *ACM. SIGCSE Bulletin*, vol. 35, no. 1, pp. 359-62, 2003.
- [45] G. A. Neuman, S. H. Wagner, and N. D. Christiansen, "The Relationship Between Work-Team Personality Composition and the Job Performance of Teams," *Group & Organization Management*, vol. 24, no. 1, pp. 28 - 45, 1999.
- [46] J. F. Pallant, *SPSS Survival Manual*, 2nd. ed.: Crows Nest, NSW, 2005.
- [47] M. A. G. Peeters, H. F. J. M. V. Tuijil, C. G. Rutte *et al.*, "Personality and Team performance: A Meta-Analysis," *European Journal of Personality*, vol. 20, pp. 377-396, 2006
- [48] A. E. Poropat, "A Meta-Analysis of the Five-Factor Model of Personality and Academic Performance," *Psychological Bulletin*, vol. 135, no. 2, pp. 322-338, 2009.
- [49] B. D. Raad, and H. C. Schouwenburg, "Personality in learning and education: Review," *European Journal of Personality*, vol. 10, pp. 303-336, July 1996, 1996.
- [50] N. Salleh, E. Mendes, and J. Grundy, "Empirical Studies of Pair Programming for CS/SE Teaching in Higher Education: A Systematic Literature Review," *IEEE Transactions on Software Engineering* (paper submitted for review), 2009.
- [51] N. Salleh, E. Mendes, J. Grundy *et al.*, "An Empirical Study of the Effects of Personality in Pair Programming using the Five-Factor Model," in *Proc. Third Int'l Symp. Empirical Software Engineering & Measurement (ESEM09)*, 2009, pp. 214 - 225.
- [52] P. Sfetsos, L. Angelis, and I. Stamelos, "Investigating the extreme programming system - an empirical study," *Empirical Software Engineering*, vol. 11, no. 2, pp. 269-301, 2006.
- [53] P. Sfetsos, I. Stamelos, L. Angelis *et al.*, "An Experimental Investigation of Personality Types Impact on Pair Effectiveness in Pair Programming," *Empirical Software Engineering*, vol. 14, pp. 187-226, 2009.
- [54] U. M. Staudinger, A. G. Maciel, J. Smith *et al.*, "What predicts wisdom-related performance? A first look at personality, intelligence, and facilitative experiential contexts," *European Journal of Personality*, vol. 12, pp. 1-17, 1998.
- [55] L. L. Werner, B. Hanks, and C. McDowel, "Pair-Programming Helps Female Computer Science Students," *Journal of Educational Resources in Computing (JERIC)*, vol. 4, no. 1, 2004.
- [56] L. A. Williams, and R. R. Kessler, "The effects of "pair-pressure" and "pair-learning" on software engineering education," *Thirteenth Conference on Software Engineering Education and Training. IEEE Comput. Soc. 2000*, pp. 59-65, 2000.
- [57] L. Williams, L. Layman, J. Osborne *et al.*, "Examining the Compatibility of Student Pair Programmers," in *AGILE 2006 Conference (AGILE'06)*, 2006.
- [58] M. J. Zyphur, J. C. Bradley, R. S. Landis *et al.*, "The Effects of Cognitive Ability and Conscientiousness on Performance Over Time: A Censored Latent Growth Model," *Human Performance*, vol. 21, pp. 1-27, 2008
- [59] B. Barry, "Composition, Process, and Performance in Self-Managed Groups: The Role of Personality," *J. of Applied Psychology*, vol. 82, no. 1, pp. 62-78, 1997.
- [60] J. A. L. Pine, "Adaptability to Changing Task Contexts: Effects of General Cognitive, Conscientiousness, and Openness to Experience," *Personnel Psychology*, vol. 53, pp. 563-593, 2000.