The Effects of Neuroticism on Pair Programming: An Empirical Study in the Higher Education Context

Norsaremah Salleh
Department of Computer Science
International Islamic University Malaysia
norsaremah@iium.edu.my

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

John Grundy
Computer Science & Software Engineering
Swinburne University of Technology, Australia
jgrundy@swin.edu.au

Giles St. J Burch
Clinical Psychology Unit
The University of Sydney
Australia
gbur9811@uni.sydney.edu.au

ABSTRACT
This paper reports on an empirical study that investigates the effects of the personality trait of neuroticism on the academic performance of students who practiced pair programming during one academic semester. The experiment was conducted at The University of Auckland involving 270 first year undergraduate students enrolled in an introductory programming course. In this study, we hypothesized that neuroticism or lack of 'emotional stability' potentially affects pair students' academic performance. However, from the analysis of our results we found lack of evidence to support this. A correlation analysis showed significant positive associations between the conscientiousness personality trait and almost all performance criteria, thus corroborating evidence reported in the educational psychology literature.

Categories and Subject Descriptors
k.3.2 [**Computer and Information Science Education**]: Computer Science Education

General Terms
Management, Experimentation, Human Factors.

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

1. INTRODUCTION
Pair programming (PP) involves teams of two people developing software where one acts as a driver, the other an observer [56]. Since its advent many educators have trialed and endorsed its use in educational settings, most often in courses focused on learning to program or improving programming skills [55], [33]. A number of factors have been explored in terms of both pair formation and influences on the pair performance during PP [27], [54].

The issue of personality in pair programming (PP) has been addressed in a number of studies (e.g. [54], [12], [49], [22], [46]) where their central theme was to investigate the impact of personality on performance of teams and individuals practicing PP. Since PP is a practice involving social interaction between two people collaboratively working closely together to solve programming and/or design problems, one can argue that its effectiveness can be potentially affected by human-related factors such as personality [21], [49].

In reviewing the literature of PP in the higher education context, we found evidence that the results of previous PP studies were inconsistent in terms of the effect or influence of personality towards pairing effectiveness [46]. These could be due to the differential set of instruments and personality frameworks used to measure personality, and the variation in the studies' context thus making it difficult to generalize the results.

The present study is an extension of our previous work [45], where we have investigated the effects of personality from the perspective of the five-factor personality model (FFM), with first year Computer Science students practicing PP in an introductory programming course. The main motivation behind that work was to look at the effect of FFM on PP, given that it had not yet been previously investigated at length particularly in teaching or academic settings.

In the present study, we focused on the neuroticism factor, which is one of the FFM's personality factors reported to have a prominent role in learning and educational context [44]. Neuroticism relates to the level of emotional stability, where high neuroticism reflects a person's negative disposition such as feeling anxiety, hostility, or self- consciousness [13]. In contrast, a person who is low in neuroticism exhibits a more resilient character represented by being composed, calm, and rarely discouraged [40]. Therefore, the aim of this study is to investigate whether or not neuroticism plays a role in differentiating the performance of students who pair programmed throughout one academic semester.

The key contribution of this paper is to add empirical evidence regarding the effects of the neuroticism personality traits towards paired students' academic performance. This will increase our understanding on the potential effects of personality towards PP's effectiveness as a pedagogical tool in Computer Science/Software Engineering (CS/SE) education. These results can also be used to better inform teachers about the implications of personalities on team and academic performance when employing PP, such that their team formation approaches are influence accordingly. Finally, we also believe it would be a useful addition to guide future research in PP team composition.

The remainder of this paper is organized as follows: Section 2 presents an overview of related work and motivation for this study. Section 3 describes briefly the five-factor personality model, followed by Section 4, where the research methodology of

ESEM'10, September 16-17, 2010, Bolzano-Bozen, Italy.

our formal experiment is described. Section 5 presents the experiment's results followed by a discussion in Section 6. Finally, Section 7 concludes our work.

2. MOTIVATION AND RELATED WORK

In more than a decade of research, researchers have investigated PP and its usefulness and effectiveness in both academic and industry settings [21], [46]. In an academic context, studies reported that PP is shown to benefits students learning outcomes in a number of ways. For example, PP is reported to help increase Computer Science students' retention rate [34], improve the course completion rate [33], increase performance in assignments [20], [35], and enhance students satisfaction and confidence level in learning programming subjects [35], [56].

These benefits however do not come without a cost. The two major issues frequently highlighted in the PP literature that hinder its effective implementation in higher education contexts are scheduling conflicts [14], [24]; and partner incompatibility [23], [31]. Such incompatibility issue might be related to psychosocial aspects such as personality and gender differences [51], [11]. Finding a compatible or matching partner is a challenge and considered a complex issue not only in academia but also in industry [50]. A recent survey by Microsoft researchers has identified "personality conflicts" as the third major problem in PP, as perceived by the developers [4]. Understanding how personality affects or relates to PP's effectiveness is therefore an important aspect that motivates us to carry out the research described herein.

In assessing personality, the Myers-Briggs Type Indicator (MBTI) has been used in most existing PP studies in academic settings [46]. Others have used the Keirsey Temperament Sorter (KTS) [50] and most recently some studies have applied the big-five or five-factor personality model (FFM) [22], [45], [47]. Although MBTI was found very popular and widely used in research in the computing and business domains, there is an emerging trend of research in CS/SE employing the FFM [22], [53], [45], [47]. The FFM, which consists of five broad personality traits (detailed in Section 3) is considered a robust taxonomy of personality and reported to receive the most support by personality traits researchers and psychologists [2], [7]. Such a growing acceptance of FFM has motivated us to employ this framework in our current study, and also in our previous work [45], [47].

Of the five PP studies [27], [28], [31], [54], [12] that investigated personality using MBTI, only one study reports that pairing works effectively for paired students with different personality types [12]. Another study by Sfetsos et al. [50], which applied KTS, also suggests that pairs consisting of heterogeneous personalities performed better than pairs with the same personality type. Other studies report no significance effects of personality in PP [27], [28], [31], [54].

To date, empirical findings using the FFM report low support for the effects of personality in PP. For instance, Hannay et al. [22] report personality as only a moderate predictor for pair performance. They suggest that the performance of pair programmers may also be affected by other factors such as expertise, and task complexity. Other empirical studiesy reported by Acuna et al. [57] investigated the relationship between personality, team processes, task characteristics, software quality and team's satisfaction in students' teams practicing the Agile XP methodology. Their findings indicate that the personality factor extraversion is positively correlated with software quality and teams with higher aggregate on agreeableness and conscientiousness achieved the highest job satisfaction [57].

Our previous work [45], [47] showed that academic performance was not significantly affected by differences in conscientiousness level among paired students; however results indicate a significant positive correlation between performance and the "openness to experience" trait. Conscientiousness was studied in our previous study because it is considered to be the most influential trait that can potentially affect academic success as well as team performance as reported in the psychology literature [8], [41], [2].

Other than the conscientiousness trait, de Raad and Schouwenberg [44] proposed that the two other significant personality traits which are educationally important and relevant for higher education are neuroticism and openness to experience [44]. In a review of personality in learning and education, they mentioned that "*particularly at the University level, highly neurotic students are probably handicapped as compared to low neurotics*." (pg. 326) [44]. Thus we believe that these traits may play a role in determining performance of students who pair programmed.

In two longitudinal studies of two British university samples, the findings showed that neuroticism is negatively and significantly related to academic performance, particularly for examination marks [9]. Similar findings were reported in their replication study [10], where the negative relationship between academic success and neuroticism was observed as a result of anxiety and impulsiveness traits. Contrary to this evidence, Komarraju et al. [30] report that "*students who are keen on performing well may also experience some degree of anxiety about being successful*" (pg. 50) [30]. Therefore, we perceived that researching aspects of neuroticism in PP seems relevant and applicable in our context, given that our aim is to help improve PP as an effective alternative pedagogical technique for CS/SE teaching.

Existing literature also report that a team consisting of low neuroticism members shows a greater ability to succeed [29]. Others also mentioned the positive correlation between 'emotional stability' and the subsequent team performance - (i.e. negative association between neuroticism and performance) [42], [2]. This is because a team consisting of highly emotionally stable members are typically characterized as self-confident and secure and hence promoting team cooperation [29], [52], [2]. Consequently, teams consisting of higher variability (heterogeneous) on neuroticism may be detrimental to performance [36].

3. THE FIVE FACTOR MODEL

The Five-Factor Model is a taxonomy of personality that is comprised of five broad personality traits - Openness to experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism [32]. 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 people who are 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 high neuroticism individual tends to be moody, anxious, nervous, and insecure [15].

In comparison with MBTI, the FFM is a trait theory of personality, based on factor-analytic studies, whereas MBTI was developed based on Jung's theory of psychological types [18]. The MBTI categorizes individual behaviour into four dimensions of personality type: *Extroversion (*E*)* vs. *Introversion* (I), *Sensing* (S) vs. *Intuition* (N), *Thinking* (T) vs. *Feeling* (F), and *Judging* (J) vs. *Perceiving* (P) [38], [39]. 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 (e.g. ENFJ) [38], while measures of the FFM sum the scores of each facet from each factor, using a 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 [44]. 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].

4. THE EXPERIMENT

The PP experiment described in this paper was conducted at The University of Auckland during semester 2, 2009. Participants were first-year students attending an introductory course for learning an object-oriented programming language in Java - Principles of Programming Course (COMPSCI 101). During this course, students learnt about basic programming concepts, and created a few small applications as part of their assignments. The formulation of hypothesis, the instruments and the experimental procedure used in this study are detailed in the following subsections.

4.1 Research Objectives

The PP experiment aimed to investigate the effect of the personality trait of neuroticism on the academic performance of students practicing PP throughout one academic semester. The rationale for using students as subjects was mainly due to the study's focus - to improve PP's effectiveness in an academic setting. The objectives of our experiment were outlined using the Goal/Question Metric template (GQM) [3] and the 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 neuroticism factors towards 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.

4.2 Hypothesis

Of the five personality constructs, neuroticism (or lack of emotional stability) is the factor that deals with feelings of anxiety, self-consciousness, impulsiveness, and vulnerability [44], [13]. Evidence suggests that neuroticism is negatively correlated with academic performance due to the effects that traits that anxiety and impulsiveness have [10]. It should however be noted that there is some evidence from organizational psychology that in certain conditions anxiety and neuroticism may actually facilitate performance [7]. On a positive side, emotional stability is consistently related to self-efficacy, which in turn, affects performance [48], [2]. Barick et al. [2] reports that teams comprising more emotionally stable members (i.e. low neuroticism) are likely to achieve higher performance when compared with teams that include even a single member who is emotionally unstable. Therefore, we posited that the level of neuroticism may influence the academic performance of students practicing PP. Therefore, we have investigated the following hypothesis in our experiment:

H_0: Differences in levels of neuroticism do not affect the academic performance of students who pair programmed.

Which is contrasted by the following alternative hypothesis:

H_A: Differences in levels of neuroticism affect the academic performance of students who pair programmed.

Table 1 shows the categorization of pairs according to students' level of neuroticism. Pair (N_{High}, N_{High}) denotes a pair combination where both students have high levels of neuroticism.

Table 1: Pair Configuration

Experimental Group	Controlled Group
Pair (N_{High}, N_{High})	Pair (N_{Low}, N_{Low})
Pair (N_{High}, N_{High})	Pair (N_{Med}, N_{Med})
Pair (N_{Med}, N_{Med})	Pair (N_{Low}, N_{Low})

Our experiment also looked into the relationship between a student's personality score with their academic performance, and their level of satisfaction and confidence when working in pairs. Academic performance was measured using tutorial exercises, assignments, a midterm test and final exam scores. Hence, academic performance was our dependant variable, and level of neuroticism (low, medium, high) our independent variable. The levels of satisfaction and confidence were measured using a questionnaire where all questions employed a five-point likert-scale.

4.3 Instrumentation and Materials

At the start of the academic semester, one of the authors provided the participants with an overview of the experiment (including PP) in one of the course lectures. During that lecture, consent forms and participant information sheets (PIS) were distributed for signing. The PIS described important information regarding the experiment and highlighted its major purpose.

Participants' personality traits were measured using a short version of the IPIP-NEO[1] because this short version had been reported to measure exactly the same traits and to also present acceptable measurement reliability as the more detailed version of

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

IPIP-NEO [25]. The personality test consists of 120 items which descriptions were authored by John A. Johnson [25]. The test produces scores in a numerical scale, with 0 and 99 representing the lowest and the highest scores for each factor, respectively. The grouping of participants per neuroticism level (see Table 2) was done based on the distribution of scores for the neuroticism trait. This was done in order to provide a more balanced number of subjects within each group.

Table 2: Personality Scores Level

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

In addition to the online personality test, participants were administered with a pre-test questionnaire to gather their demographic information as well as their programming competency level. In addition, short questionnaires were given to the students during each of the PP sessions in order to measure their satisfaction and confidence level working with their partner.

Inline with the University's requirements, we have obtained the approval of the University of Auckland's Human Participants Ethics Committee for this study prior to performing the data collection.

4.4 Experimental Setup

The experiment took place during the compulsory weekly tutorial sessions of the COMPSCI 101 course, run by a tutor and a few Teaching Assistants (TAs). We followed the same procedure carried out in our previous study [47], where each of the tutorial sessions was treated as an independent formal experiment. Students' personality and demographic data were gathered at the start of the semester. An online version of the IPIP-NEO was used to measure students' personality against the FFM. The results of the personality profiling were then used to allocate partners. For this purpose, the scores on the neuroticism trait were used to assign paired students in three possible groups, representing three different levels of neuroticism: low, medium and high.

In every tutorial, pairs were allocated randomly within each group. Thus a "single-factor between-group design" [37] was the research design employed in this study. Every tutorial lasted for two hours where the first 45 minutes were used by the tutor to explain the topic, and the remaining 75 minutes were allocated for students to solve the exercises in pairs. To allow for "pair-jelling" [55], students worked with their partner for an initial period of 30 minutes; and then they were required to swap their roles.

Before the end of each tutorial, students provided feedback about working with their partner by filling out a short questionnaire. Students indicated their satisfaction level of working with their partner by answering a question "*Please rate how satisfied are you working with your partner*" on a scale from 1 to 5 (very dissatisfied to very satisfied). To measure their confidence level, students were asked a question "*How do you rate your level of confidence solving the exercises with your partner?*" also using the scale from 1 – 5 (very low to very high). The exercises given during the tutorials were graded, thus contributing towards their final grade. In addition, assignments and test were also graded, however completed individually.

The outcomes measured from the experiment were determined by the scores on the tutorial exercises, three assignments, a mid-term test and final exam. During the tutorial sessions, students were required to solve a minimum two programming problems with the assigned partner. Since tutorials varied from week to week, the experiments were designed in such a way as to minimize the effect of confounding factors due to differences in the tasks and level of complexity of the exercises assigned to the students. Therefore, the tasks and exercises remained the same throughout each week.

5. RESULTS

5.1 Subjects

A total of 270 students were enrolled in the course. Of these, 202 were male students (74.8%), and subjects' age ranged from 19 to 47 years old (the mode age = 19 years). Of the 81 students who answered the demographics questionnaire, 64 students (79%) did not have any previous work experience. Only 77 out of 270 students (29%) declared Computer Science as their major. Of the 270 students, 118 students (44%) completed the personality test and have taken either the mid-term test or final exam. Therefore, the sample size used in our analysis was of 118 students.

The boxplot in Figure 1 shows the distribution of personality data for all the five traits. The box represents the middle 50% of the scores, with the upper and the lower tails indicating the 75th and 25th percentiles, respectively. The line drawn across the box shows the median value. Figure 1 shows that the median scores for neuroticism and agreeableness are considerably higher than those for the other factors, whereas the median score for openness is the lowest. The outliers represent cases where the openness trait score is very high (i.e. above 74).

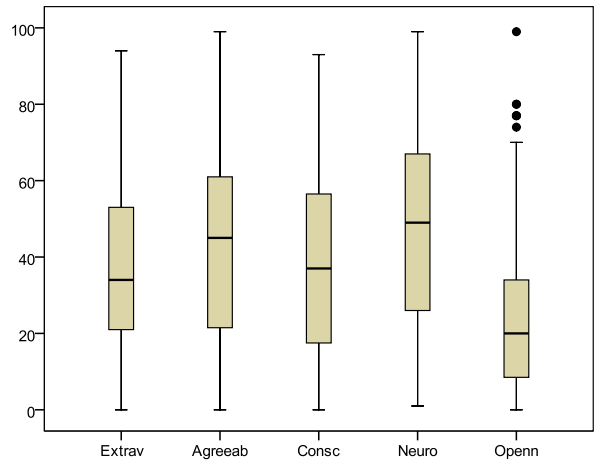

Figure 1: Distribution of FFM scores

5.2 Correlational Analysis

Figure 2 shows the distribution of tutorial scores for each of the neuroticism levels. All boxplots showed similar medians, where the flattest distribution belonged to the medium neuroticism group, followed by the low neuroticism group. The outliers present in the low neuroticism group represent students who did not regularly attend the tutorial sessions thereby missing some of the exercises; outliers for the high neuroticism group represents 22% of the students who obtained marks lower than 9.0. This

indicates students in the high neuroticism group presented higher scores in tutorial exercises than their counterparts.

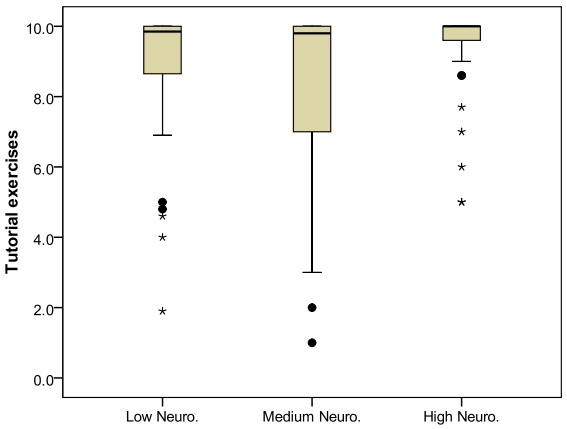

Figure 2: Comparison of tutorial scores between groups

Figure 3 shows the boxplot of assignment scores by neuroticism level. The median for the high neuroticism group was higher than the other two, followed by the low neuroticism group. The maximum possible scores for the tutorial and assignments were 10 and 15 respectively.

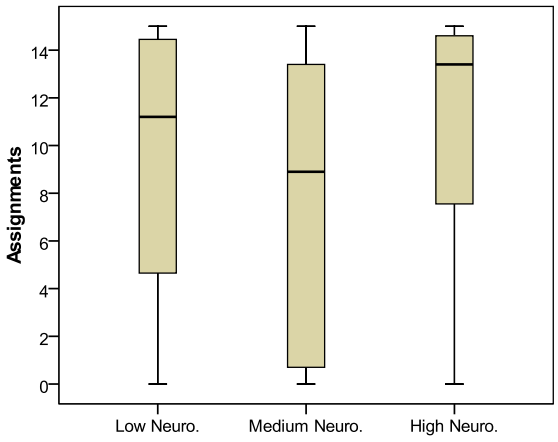

Figure 3: Comparison of assignments scores between groups

In terms of the midterm test scores, the median for the high neuroticism group was slightly higher than the other two (see Figure 4). In addition, the flattest distribution was observed for the medium neuroticism group, suggesting this group was more heterogeneous compared to the other two groups. The outliers in this boxplot showed cases where the midterm score was lower than 20%.

A similar pattern was observed regarding the medians of final exam scores (see Figure 5). However, in this case, all three distributions have similar dispersion, and there are no outliers. The maximum possible scores for mid-term and final exam were both 100.

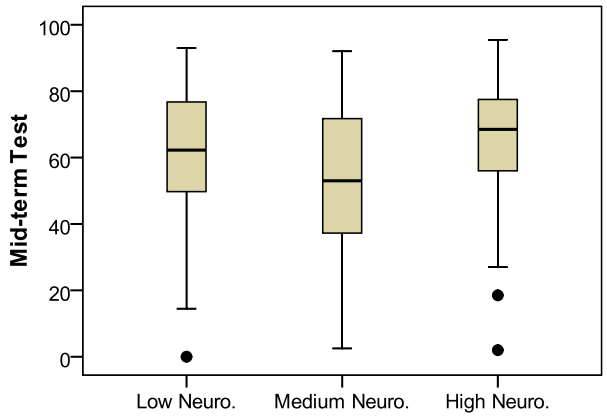

Figure 4: Comparison of mid-term scores between groups

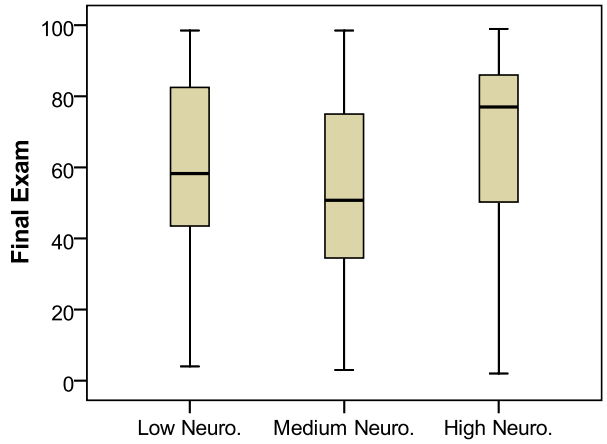

Figure 5: Comparison of exam scores between groups

In order to measure the strength of association between neuroticism levels and academic scores we employed the Pearson's correlation coefficient (α = 0.05) (see Table 3). A parametric test was chosen because the sample size used in our study was not considered small [41]. No statistically significant association was found between neuroticism and any measure of academic performance. As Table 3 shows, we also measured the level of association between the other four personality traits and students' academic performance. These results showed a statistically significant correlation between participants' conscientiousness levels and performance in tutorial, assignments, and test scores: $r(118) = 0.29$, $p<0.01$ for tutorials, $r(116) = 0.19$, $p<0.05$ for assignments, and $r(116) = 0.19$, $p< 0.05$ for test.

The positive significant correlation between conscientiousness and assignments was consistent with those from our previous study [45]. The findings regarding conscientiousness were also corroborated with those reported in the psychology literature (e.g. [10], [8], [19]) which suggest conscientiousness as a consistent predictor for academic performance due to the nature characteristics of highly conscientious individuals (i.e. persisting, achieving, grades orientation etc.)

Table 3: Correlation between traits and academic performance

	1	2	3	4	E	A	C	N
1	1							
2	0.56**	1						
3	0.36**	0.54**	1					
4	0.46**	0.68**	0.83**	1				
E	0.00	-0.01	0.05	-0.02	1			
A	-0.00	-0.02	0.09	-0.02	0.09	1		
C	0.29**	0.19*	0.19*	0.15	0.28**	0.21*	1	
N	0.07	0.05	0.03	0.01	-0.24*	-0.15	-0.25**	1
O	0.02	0.01	0.18	0.15	0.07	0.24**	0.01	0.21*

1. Tutorial Scores 2. Assignments 3. Test 4. Exam
(E) Extraversion (A) Agreeableness
(C) Conscientiousness (N) Neuroticism (O) Openness
**. Correlation is significant at the 0.01 level (two-tailed).
*. Correlation is significant at the 0.05 level (two-tailed).

5.3 Analysis of the Hypothesis Testing

The hypothesis investigated herein was tested using the one-way between-group analysis of variance (ANOVA) ($\alpha = 0.05$) in order to compare the effects of three levels of neuroticism on paired students' academic performance. Table 4 shows the values for the mean scores and standard deviations for each neuroticism level. The mean values for all neuroticism levels were quite similar for tutorial scores, however this trend was not observed for the other performance measures.

Table 4: Means and Standard Deviations for Neuroticism Levels per Performance Measure

Performance Measures	Neuroticism Levels	N	Mean	SD
Tutorials (range: 0 to 10)	Low Neuro.	40	8.75	2.03
	Medium Neuro.	45	8.11	2.65
	High Neuro.	33	9.27	1.46
	Total	118	8.65	2.19
Assignments (range: 0 to 15)	Low Neuro.	40	10.14	5.13
	Medium Neuro.	45	8.21	5.65
	High Neuro.	33	10.98	4.59
	Total	118	9.64	5.28
Test (range: 0 to 100)	Low Neuro.	40	59.85	21.57
	Medium Neuro.	43	52.20	23.71
	High Neuro.	33	64.00	21.92
	Total	116	58.19	22.82
Final Exam (range: 0 to 100)	Low Neuro.	38	59.97	23.59
	Medium Neuro.	42	52.59	26.75
	High Neuro.	31	63.06	31.32
	Total	111	58.05	27.22

The ANOVA results (see Table 5) showed that at the $p < 0.05$ level there was no statistically significant difference in academic performance between the three groups of neuroticism (i.e. $F(2,115)=2.8$, $p=0.07$, for tutorials; $F(2,115)=3.0$, $p=0.054$, for assignments; $F(2,113)=2.74$, $p=0.07$, for midterm test; $F(2,108)=1.48$, $p=0.23$, 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. Therefore, based on our data, we found that paired students academic performance was not significantly affected by differences in neuroticism levels.

Table 5: ANOVA Results

		SS	df	MS	F	*p*
Tutorials	BW	26.16	2	13.08	2.80	0.07
	WG	537.05	115	4.67		
	Total	563.21	117			
Assign.	BG	161.96	2	80.98	3.00	0.05
	WG	3106.47	115	27.01		
	Total	3268.43	117			
Test	BG	2768.22	2	1384.11	2.74	0.07
	WG	57133.67	113	505.61		
	Total	59901.89	115			
Exam	BG	2169.81	2	1084.91	1.48	0.23
	WG	79357..96	108	734.79		
	Total	81527.78	110			

SS – Sum of Squares MS – Mean Squares
BG – Between Groups WG – Within Goups

5.4 Results on Satisfaction and Confidence

We gathered data on paired students' satisfaction and confidence working with their partner using a questionnaire distributed during the tutorial sessions. These data were gathered for eight weeks. We did not gather the data for the first two weeks in order to give students ample time to familiarize themselves with PP. Data were analyzed separately as each tutorial was treated as a single independent "mini-experiment". The dependent variable satisfaction was measured on a scale from 0 (*very dissatisfied*) to 5 (*very satisfied*); and confidence level was measured on a scale from 0 (very low) to 5 (very high).

The response rate of the post-experimental questionnaire was approximately 42% for every tutorial. On average, 60 (84%) out of an average of 72 students attending the tutorials, were satisfied working with their partner. Table 6 shows the mean rank for the Kruskal-Wallis test ($\alpha = 0.05$) which was used to compare the satisfaction rates between the neuroticism levels. The group with the higher rank indicates the higher rates for satisfaction. Overall results demonstrated that the satisfaction levels of paired students were not affected by different levels of neuroticism. Of eight weeks of tutorials, only the last tutorial (Tutorial 10) showed a significant difference in satisfaction across the three levels of neuroticism: χ^2 (2, 68)=13.12, p=0.001. Nevertheless, these data also show the trend that, according to their mean rank, paired students with low neuroticism had higher satisfaction compared with the other two neuroticism levels.

Table 6: Mean Rank for Satisfaction Level

	Neuro. Level	N	Mean Rank	Sig.	Satisfied/Very Satisfied (%)
Tut. 3 N=46	Low	16	25.00	0.22	82.6
	Medium	20	25.30		
	High	10	17.50		
Tut. 4 N=95	Low	29	51.60	0.48	83.2
	Medium	35	48.67		
	High	31	43.87		
Tut. 5 N=63	Low	18	34.61	0.45	90.5
	Medium	28	29.11		

	High	17	34.00		
Tut. 6 N=65	Low	23	37.22	0.22	87.7
	Medium	24	32.94		
	High	18	27.69		
Tut. 7 N=71	Low	25	32.84	0.54	94.4
	Medium	23	38.00		
	High	23	37.43		
Tut. 8 N=54	Low	15	31.10	0.49	87.0
	Medium	20	26.80		
	High	19	25.39		
Tut. 9 N=69	Low	25	37.80	0.46	91.3
	Medium	27	34.94		
	High	17	30.97		
Tut.10 N=68	Low	20	45.30	0.00	82.4
	Medium	18	36.06		
	High	30	26.37		

Table 7 shows the mean rank for paired students' confidence level based on the analysis of the returned surveys. There was only one tutorial that presented a significant difference of confidence level across the three groups (Tutorial 4): χ^2 (2, 95)=10.69, p=0.005. This particular result indicates that the low neuroticism group obtained the highest confidence level compared with others. Overall, we found that confidence in solving the exercises was generally high among the low and medium neuroticism groups. These results suggest the tendency of students of lower or moderate neuroticism to believe in the correctness of their programming solutions compared to the high neuroticism pairs.

Table 7: Mean Rank for Confidence Level

	Neuro. Level	N	Mean Rank	Sig.	% of high confidence
Tut. 3 N=46	Low	16	25.19	0.12	78.3
	Medium	20	25.80		
	High	10	16.20		
Tut. 4 N=95	Low	29	56.86	0.01	81.1
	Medium	35	51.14		
	High	31	36.16		
Tut. 5 N=63	Low	18	35.58	0.46	82.5
	Medium	28	29.32		
	High	17	32.62		
Tut. 6 N=65	Low	23	38.13	0.29	83.3
	Medium	24	31.77		
	High	18	30.08		
Tut. 7 N=71	Low	25	33.76	0.62	91.6
	Medium	23	38.89		
	High	23	35.54		
Tut. 8 N=54	Low	15	29.97	0.57	88.9
	Medium	20	28.10		
	High	19	24.92		
Tut. 9 N=69	Low	25	33.54	0.74	86.9
	Medium	27	37.06		
	High	17	33.88		
Tut.10 N=68	Low	20	41.20	0.09	85.3
	Medium	18	35.06		
	High	30	29.70		

6. DISCUSSION

Although neuroticism is reported to be related with the tendency to have "poorer" performance in some studies (e.g. [2], [10], [5]) the findings from our study do not support this view. Based on the ANOVA analysis, we did not find any significant difference in performance between paired students of different neuroticism levels. These results are consistent with other findings from previous research linking neuroticism to academic performance among students in tertiary institutions (e.g. [19], [8], [17]).

In regard to the relationship between personality and team performance, a meta-analysis by Peteers et al. [42] suggests that the elevation in emotional stability (i.e. low neuroticism) is not significantly related to team performance due to the "broad concept" or wider impression represented by this trait. Instead, they proposed that the facets within the neuroticism trait (e.g. self-consciousness, impulsiveness) should be empirically tested in order to obtain a more genuine effect [42].

Existing research evidence also suggests the potential of moderator effects which could possibly influence the personality-team relationship [5], [42]. One of such effects is the type or the complexity of the task engaged by the team [42], [6]. Bowers et al. [6] suggest that the personality homogeneity of team members had very little effects on team performances, particularly on low-difficulty tasks [6]. Thus, the lack of statistically insignificant findings in our study is probably related to the less complex tasks assigned to students. Future research should investigate the mediator variable in order to better understand the impact of personality traits on performance. For example, a qualitative study on the nature of collaboration in PP by Walle & Hannay [53] revealed some relationships between personality traits and the type of collaborations that may affect pair performance [53].

The positive significant correlation between conscientiousness and performance in tutorials, assignments, and midterm test demonstrated in our findings indicate the tendency that students' conscientiousness level may play a role in predicting performance. This is because conscientiousness is the one consistently bearing significant positive relationship with high achievement of academic as well as team performance as also reported in the literature [2], [16], [42], [43]. We also observed a similar correlation when investigating conscientiousness in our previous work [45]. Therefore we believe that future studies need to investigate further the conscientiousness trait in order to assess whether findings converge.

In terms of the satisfaction level, overall results showed that differences in neuroticism levels were not significant in affecting students' contentment while working in pairs. Despite these results, lower neuroticism pairs scored higher satisfaction in most tutorials compared with the other groups (see Table 7). This perhaps relates to the common characteristic of low neuroticism individuals being well adjusted people and likely to excel in team settings, as reported by Driskel et al. [15].

6.1 Threats to the Validity

One of the potential threats to the internal validity of our study relates to the issue of changing partners during the tutorial. Some students failed to turn up to their allocated tutorial and attended a different session without informing the tutor. This could have led to students being paired with students from different neuroticism groups. However, according to the tutor, these uncontrolled

circumstances occurred sparingly thus minimizing the potential to bias the results.

Another potential threat relates to gender differences. As reported by Schmitt [48], the interaction of neuroticism and gender had significant impact on self-efficacy and performance. In our study, approximately 75% of the subjects are male students; therefore we believe that the probability of such significant impact would be minimal due to the lower number of females enrolled in the course.

Another limitation refers to the fact that the performance measures used in this study may also be affected by levels of cognitive ability. In this study we used students' academic performance as surrogate measures of PP's effectiveness. Thus, there is a possibility that performance is affected by students' ability and competency in programming. However, since the study aimed to improve students' learning due to practicing PP throughout the entire semester, measuring their academic performance is in our view appropriate to our context. In addition, empirical evidence shows that the predictive power of one's cognitive ability in association with academic performance is relatively low compared to personality traits [19]. Therefore, students' cognitive ability may not have affected the results presented herein.

7. CONCLUSIONS AND FUTURE WORK

In conclusion, the work presented herein extends our previous work on understanding the effects of personality traits on PP's effectiveness, where effectiveness is measured as students' academic performance. The findings showed that paired students performance was not significantly affected by the different levels of neuroticism for the sample employed in this study. The lack of support for the alternative hypothesis could be attributed to the low complexity of tasks assigned to students, and perhaps the existence of moderator variables mediating the relationship between personality traits and performance.

The positive significant association shown between conscientiousness in almost all performance measures also warrants further investigation. This is because, out of the five personality factors, conscientiousness is reported to be the most significant predictors of academic performance for tertiary level students [10], [9]. In addition, team members consisting of highly and similarly on conscientiousness are reported to achieve better performance [42]. The findings from the present study also indicate that students' satisfaction and confidence level did not differ depending on the levels of neuroticism when pairing.

Preliminary advice for educators looking to employ PP for introductory programming tasks, based on our experimental findings, is that different levels of neuroticism do not appear to significantly impact academic performance when engaging in PP. Thus, PP group formation and monitoring may be able to ignore the neuroticism levels of different students for such introductory programming tasks. The level of satisfaction with PP among pairs with high, medium and low neuroticism is broadly consistent, though lower neuroticism pairs tend to be a little more satisfied with their PP experiences. However, conscientiousness does appear to significantly associated with performance, but empirical evidence is needed to confirm the cause-and effect of this relationship.

We believe that an interesting direction for future work relates to exploring whether PP mitigates neuroticism, at least for students engaging in PP tasks. For instance, if high neuroticism students work together, or if they are paired with lower neuroticism individuals, would the pair work facilitate students to better cope with anxiety or other negative aspects of neuroticism? These questions are open for future investigation. In the future, it would also be useful to carry out replication studies of this study in order to help increase the credibility of the results, and to facilitate the generation of new knowledge [26].

8. ACKNOWLEDGMENTS

The study was funded by the Ministry of Higher Education Malaysia. The author would like to thank Ann Cameron, Adriana, and Karen for the help given to run the experiment. Thanks also to CS101 demonstrators for their assistance during the tutorials, and to all students 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, and M.K. Mount, "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] A. Begel, and N. Nagappan, "Pair Programming: What's in it for me?," in ESEM'08, Kaiserslautern, Germany, 2008, pp. 120-128.

[5] S. T. Bell, "Deep-Level Comparison Variables as Predictors of Team Performance: A Meta-Analysis," *J. Applied Psychology,* vol. 92, no. 3, pp. 595-615, 2007.

[6] C. A. Bowers, J. A. Pharmer, and E. Salas, "When Member Homogeneity is Needed in Work Teams: A Meta-Analysis," *Small Group Research,* vol. 31, no. 3, pp. 305 - 327, 2000.

[7] G. 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, and C. Hamaker, "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 predicts academic performance: Evidence from two longitudinal University samples," *J. of Research in Personality,* vol. 25, pp. 319-338, 2003.

[10] . T. Chamorro-Premuzic, and A. Furnham, "Personality Traits and Academic Examination Performance," *European Journal of Personality,* vol. 17, pp. 237-250, 2003.

[11] K.S. Choi, “A Discovery and Analysis of Influencing Factors of Pair Programming,”. Unpublished Ph.D. Dissertation, New Jersey Institute of Technology, USA, 2004.

[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] P. T. Costa, and R. R. McCrae, “Domain and facets: Hierarchical personality assessment using the Revised NEO personality inventory,” *J. of Personality Assessment,* vol. 64, pp. 21-50, 1995.

[14] T. H. DeClue, “Pair Programming and Pair trading:Effects on Learning and motivation in a CS2 courses,” *Journal of Computing Sciences in Colleges,* vol. 18, no. 5, pp. 49-56, 2003.

[15] J. E. Driskell, E. Salas, F. F. Goodwin, and P.G. O’Shea, “What Makes a Good Team Player? Personality and Team Effectiveness,” *Group Dynamics: Theory, Research, and Practice,* vol. 10, no. 4, pp. 249-271, 2006.

[16] A. English, R. L. Griffith, and L. A. Steelman, “Team Performance: The Effect of Team Conscientiousness and Task Type,” *Small Group Research,* vol. 35, no. 6, pp. 643-665, 2004.

[17] 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.

[18] 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.

[19] A. Furnham, T. Chamorro-Premuzic, and F. McDougall, “Personality, cognitive ability, and beliefs about intelligence as predictors of academic performance,” *Learning and Individual Differences,* vol. 14, pp. 49-66, 2003.

[20] B. Hanks, C. McDowell, D. Draper, and M. Krnjajic, “Program quality with pair programming in CS1,” *ACM. SIGCSE Bulletin,* vol. 36, no. 3, pp. 176-80, 2004.

[21] J. E. Hannay, T. Dyba, E. Arisholm, and D.I.K. Sjoberg , “The Effectiveness of Pair Programming: A Meta-Analysis,” *Information and Software Technology,* vol. 51, pp. 1110-1122, 2009.

[22] J. E. Hannay, E. Arisholm, H. Engvik, and D.I.K. Sjoberg “Effects of Personality on Pair Programming,” *IEEE Transaction on Software Engineering,* vol. 36, no. 1, pp. 61-80, 2010.

[23] C.-w. Ho, *Examining Impact of Pair Programming on Female Students,* TR-2004-20, North Carolina State University, Raleigh, NC, 2004.

[24] E. V. Howard, “Attitudes on using pair-programming,” *Journal of Educational Technology Systems,* vol. 35, no. 1, pp. 89-103, 2006.

[25] J. Johnson. "The IPIP-NEO Personality Assessment Tools," accessed July 2008; http://www.personal.psu.edu/j5j/IPIP/.

[26] N. Juristo, and S. Vegas, “Using Differences among Replications of Software Engineering Experiments to Gain Knowledge,” in ESEM'09, Lake Buena Visa, Florida, 2009, pp. 356-366.

[27] N. Katira, L. Williams, E. Wiebe, C. Miller, S. Balik, and E. Gehringer, “On understanding compatibility of student pair programmers,” *ACM. SIGCSE Bulletin,* vol. 36, no. 1, pp. 7-11, 2004.

[28] 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.

[29] 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.

[30] M. Komarraju, S. J. Karau, and R. R. Schmeck, “Role of the Big Five personality traits in predicting college students' academic motivation and achievement,” *Learning and Individual Differences,* vol. 19, pp. 47-52, 2009.

[31] 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.

[32] 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.

[33] C. McDowell, L. Werner, H. Bullock, and J. Fernald, “The effects of pair-programming on performance in an introductory programming course,” in ACM. SIGCSE Bulletin, 2002, pp. 38-42.

[34] C. McDowell, L. Werner, H. E. Bullock, and J. Fernald, “Pair programming improves student retention, confidence, and program quality,” *Communications of the ACM,* vol. 49, no. 8, pp. 90-5, 2006.

[35] E. Mendes, L. B. Al-Fakhri, and A. Luxton-Reilly, "Investigating pair-programming in a 2nd-year software development and design computer science course." pp. 296-300.

[36] S. Mohammed, and L. C. Angell, “Personality Heterogeneity in Teams: Which Differences Make a Difference for Team Performance?,” *Small Group Research,* vol. 34, no. 6, pp. 651 - 677, 2003.

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

[38] 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 edition ed., 1998.

[39] I. B. Myers, and P. Myers, *Gifts Differing: Understanding Personality Type*: Davies-Black Publishing, 1995.

[40] M. Ogot, and G. E. Okudan, “The Five-Factor Model Personality Assessment for Improved Student Design Team Performance,” *European Journal of Personality,* vol. 31, no. 5, pp. 517-529, October 2006, 2006.

[41] J. F. Pallant, *SPSS Survival Manual*, 2nd. ed.: Crows Nest, NSW, 2005.

[42] 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,* vol. 20, pp. 377-396, 2006.

[43] 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.

[44] 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.

[45] N. Salleh, E. Mendes, J. Grundy, and G.S.J. Burch, "An Empirical Study of the Effects of Personality in Pair Programming using the Five-Factor Model," in ESEM09, Lake Buena Vista, Florida, 2009.

[46] 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), 2010.

[47] N. Salleh, E. Mende, J. Grundy, and G.S.J. Burch, "An empirical study of the effects of conscientiousness in pair programming using the five-factor personality model," in ICSE 2010, Cape Town, SA, 2010.

[48] N. Schmitt, "The interaction of neuroticism and gender and its impact on self-efficacy and performance," *Human Performance,* vol. 21, pp. 49-61, 2008.

[49] 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," *Extreme Programming and Agile Processes in Software Engineering. 7th International Conference, XP 2006. Proceeding (Lecture Notes in Computer Science Vol.4044). Springer-Verlag. 2006,* vol. 4044, pp. 43-52, 2006.

[50] P. Sfetsos, I. Stamelos, L. Angelis, and I. Deligiannis, "An Experimental Investigation of Personality Types Impact on Pair Effectiveness in Pair Programming," *Empirical Software Engineering,* vol. 14, pp. 187-226, 2009.

[51] T. VanDeGrift, "Coupling pair programming and writing: learning about students' perceptions and processes," in ACM. SIGCSE Bulletin, pp. 2-6, 2004.

[52] A. E. M. v. Vianen, and C. K. W. D. Dreu, "Personality in Teams: Its relationship to social cohesion, task cohesion, and team performance," *European Journal of Work and Organizational Psychology,* vol. 10, no. 2, pp. 97-120, 2001.

[53] T. Walle, and J. E. Hannay, "Personality and the Nature of Collaboration in Pair Programming," in ESEM 09, Florida, 2009, pp. 203-213.

[54] L. Williams, L. Layman, J. Osborne, and N. Katira, "Examining the Compatibility of Student Pair Programmers," in AGILE 2006 Conference (AGILE'06), 2006.

[55] L. Williams, R. R. Kessler, W. Cunningham, and R. Jeffries, "Strengthening the case for pair programming," *IEEE Software,* vol. 17, no. 4, pp. 19-25, 2000.

[56] L. Williams, C. McDowell, N. Nagappan, J. Fernald, and L. Werner, "Building pair programming knowledge through a family of experiments," *Proceedings 2003 International Symposium on Empirical Software Engineering. ISESE 2003. IEEE Comput. Soc. 2003*, pp. 143-52, 2003.

[57] S. T. Acuna, M. Gomez, and N. Juristo, "How do personality, team processes, and task characteristics relate to job satisfaction and software quality?," *Information and Software Technology,* vol. 51, no. 3, pp. 627-639, 2009.