
Volume 25, Issue 4, July 2009
ISSN 0747-5632
ELSEVIER
COMPUTERS IN HUMAN BEHAVIOR
Robert D. Tennyson, Editor
Paul Kirschner, Associate Editor
Jen-Her Wu, Associate Editor
Including the Special Issue
The Use of Support Devices in Electronic Learning Environments
Guest Editors
Geraldine Clarebout, Holger Horz and Jan Elen

Computers in Human Behavior 25 (2009) 844–852

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Pair dynamics in team collaboration

Kyungsub Steve Choi [a,*], Fadi P. Deek [b], Il Im [c]

[a] Computer Information Systems Department, Manhattan College, Riverdale, NY 10471, USA
[b] Faculty of Information Systems, College of Computing Sciences, New Jersey Institute of Technology, 323 M.L.King Boulevard, University Heights, Newark, NJ 07102-1982, USA
[c] Faculty of Information Systems Department, School of Business, Yonsei University, 134 Shinchon-Dong, Seodaemoon-Ku, Seoul 120-749, Republic of Korea

ARTICLE INFO

Article history:
Available online 15 May 2008

Keywords:
Pair dynamics
Pair programming
Team programming
Extreme programming

ABSTRACT

From a recent field survey with a group of professional pair programmers, it was revealed that the programmers perceive a partner's personality, cognitive style and communication skills as the top three factors that lead to prudent pair programming. Based on this finding, the factors personality and communication skills, along with gender were selected for an experiment to analyze if a partner's human, intrinsic values influence the pair programming experience, specifically in the levels of satisfaction, compatibility, communication, and confidence. A total of 128 students majoring in Management Information Systems, Information Systems, and Information Technology participated in the experiment. Of the 68 undergraduates, 40 were first-year students and 28 were juniors; the remaining 60 were Master's degree graduate students. The students were formed into a total of 64 pairs based on their personality, level of communication skills, and gender. A total of three visits were made. During the first two visits, a set of four programming problems was used in four programming sessions lasting 45 min each; two were individual programming sessions and two were pair programming sessions. At the end of each visit, a questionnaire was administered and collected. The questionnaire results revealed that the various Myers–Briggs Type Indicator (MBTI) personality combinations did not significantly influence the levels of communication, satisfaction, confidence, and compatibility. The pairs that exhibited a high level of communication between partners did not necessarily experience a high level of satisfaction or exhibit compatibility between partners, nor did they have a high level of confidence regarding the finished product. The communication skill level seemed to have an impact on communication only. Similar to many previous gender-focused literatures, the same gender pairs did exhibit significantly higher levels of communication, satisfaction and compatibility than the mixed gender pairs. Within the same gender pairs, the female–female pairs showed a much higher level than the male–male pairs in those categories. Contrariwise, the same gender pairs did not show a significantly higher confidence level than the mixed gender pairs about their finished product.

1. Motive

In accordance with the standard business practices of corporate America, it has become a much more familiar sight in college classrooms for instructors to give group projects, rather than individual projects. This practice teaches the students to become more of a team player in the business world (Maznevski & Distefano, 2000). Moreover, this pedagogical practice teaches the students how to compromise and also respect the opinions of other students (Gallavan & Juliano, 2007). On any given team project, aside from the course-specific information and required task, there is bound to be some degree of cognitive collaboration (Winter & Joan C.; Waner, 2005; Amason & Sapienza, 1997; Schellens & Valcke, 2005). In this paper we define 'cognitive collaboration' from an intra-group perspective, as a process where the group members are productively sharing ideas, discussing thoughts, reasoning logics, analyzing concepts, and evaluating views to achieve the group's common goal.

Among various applications of cognition collaboration, a good example is working on a computer programming task where students collaborate in completing the task. The strenuous nature of computer programming demands a high level of concentration and focus, as well as a distinct problem solving style and cognition (Mayer, 1981). As each person is different in his or her approach in understanding and analyzing the problem, devising a solution plan, applying the tools, and executing the solution plan, these cognitive behaviors will be manifest through each person's computer code design. Using different computer commands, algorithms, scripts, shorter code lines, or longer code lines, are all reflective of how a person is different and unique in his or her cognitive process. Thus,

* Corresponding author. Tel.: +1 718 862 7309; fax: +1 718 862 7440.
E-mail addresses: kyungsub.choi@manhattna.edu (K.S. Choi), fadi.deek@njit.edu (F.P. Deek), il.im@yonsei.ac.kr (I. Im).

0747-5632/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.chb.2008.09.005

the psychology of computer programming manifests one's individualism and cognitive attributes (Weinberg, 1998).

Consequently, this phenomenon would be amplified if two individuals with different cognitive styles are forced to collaborate and compromise on writing computer codes by sharing a personal computer with one keyboard and one monitor. This programming paradigm is called pair programming (Beck, 2000; Williams & Kessler, 2002). Pair programming (PP) is defined as a programming activity where two individuals sit next to each other and share one keyboard and display. This can be done virtually where two programmers, geographically located in different places, share code development in real-time and take turns coding (Stotts et al., 2003). The one who is coding is called the "driver" and the one who is not coding is called the "navigator". The "navigator" reviews code and monitors the "driver's" coding process. Each programmer takes a turn in being the "driver" and the "navigator." While coding, the two actively collaborate in designing, coding, and reviewing. Thus, PP eliminates the overhead cost of a more formal code review.

This context of being "harnessed to each other" produces explicit and implicit expectations of friendly and hostile compromises (Williams & Kessler, 2000; Williams & Kessler, 2002; Winter et al., 2005). In order to shed more light on this research question, a preliminary field survey was taken from a group of professional programmers (Choi & Management., 2007). The survey asked the participants to choose factor(s) they perceived to influence PP, and the most popular factor was personality, followed by cognitive style and communication skills. The survey participants had an average of 15 years of professional programming experience and an average of 21 months of PP. These findings imply that a positive PP experience, where the partners feel satisfaction, compatibility, communication, and confidence, hinges largely on intrinsic human values and how the two programmers interact and share those values with each other. Under the PP context and by selectively manipulating personal attributes, this particular study investigates the level of two programmers' perceived group communication, perceived group satisfaction, perceived confidence in group output (computer codes), and perceived group compatibility.

2. Literature review

Instead of viewing computer programming as a single task, one may view it as an assembly of multiple phases consisting of problem analysis, code design, coding, testing, debugging, and documentation. The logic behind this proposal is that each phase requires different abilities and instincts. Hence, some claim that a certain cognitive or personality type performs better than other types in certain phases (Bishop-Clark, 1995; Weinberg, 1998; Da Cunha & Greathead, 2004; Yourdon, 1997). This may explain why so many "hard core" coders detest lengthy white comments and code development documentations.

Cheng, Luckett, and Schulz, 2003, using Myers–Briggs Type Indicator (MBTI), studied the business decision quality variation on different cognitive styles of a group. The results showed that cognitively diverse pairs performed significantly better than homogeneous pairs. This synchronizes with the statement made by Myers and Myers (1995) that "the shared preference gives a team common ground and the team's dissimilar preference gives them, as a team, a wider range of expertness than either has alone". Bradley and Hebert (1997) studied the productivity difference between two teams where each team included different personality types. The study proposed that IS development team performance is at least partially related to the team's personality-type composition.

Katira et al., 2004 studied the PP experience and compatibility of under-represented groups, specifically female and minority students, in computer science. The results revealed that students were more compatible with partners whose programming skill levels were similar. Pairing two female students was likely to result in a compatible pair in the undergraduate classroom, while mixing gender pairs was less likely to lead to compatibility. Thomas, Ratcliffe, and Robertson, 2003 showed that a group of students who had prior programming experience did not enjoy PP as much as others with no programming experience. The students produced their best work when paired with students with similar self-confidence levels. Asides from personal attributes, negotiating and communication skill may also play a role when members have to compromise and collaborate (Thompson & Hastie, 1990).

3. Setting up the experiment

Based on the field survey result, the factors personality, level of communication skill and gender were selected for this experiment's independent variables. The gender factor was added to observe if it revealed similar results as many previous gender-focused experiments (Stewart, Cooper, Stewart, & Friedley, 1996; Tannen, 1991; Savicki, Kelley, & Lingenfelter, 1996; Tor, 1996), which reported stronger affinity among the same gender group than the mixed gender group. For this study, a total of 128 students majoring in various computing majors started the experiment. Of the 68 undergraduates, 40 were first-year students and 28 were juniors; the remaining 60 were Master's degree graduate students. Majors included Management Information Systems, Information Systems, and Information Technology. The participants were profiled by their personality, the level of communication skill, and gender in the following manner.

For the personality profile, Myers–Briggs Type Indicator (MBTI) was used. MBTI (Bayne, 1995; Keirsey, 1998; Myers & Myers, 1995) is used in 84 of the Fortune 100 companies and in more than 50 countries according to CCP Inc., the official MBTI material distributor (CCP website 1) MBTI has four major categories, and each category presents two polar opposites. They are: (1) where a person focuses his or her attention – extraversion (E) and introversion (I), (2) the way a person gathers information – sensing (S) and intuition (N), (3) the way a person makes decisions – thinking (T) and feeling (F), and (4) how a person deals with the outer world – judging (J) and perceiving (P). For example, a person could be MBTI profiled with INFJ. This means the person tends towards introversion, intuition, feeling, and judging.

Type theory (Bayne, 1995) generally asserts that one of the four preferences – sensing, intuition, thinking or feeling – usually dominates the others. For example, a person uses the dominant type the most and feels most comfortable when using it; it is an essential part of the person at his or her best. A person uses the dominant type in his or her daily life. Complementing the dominant type is the auxiliary or secondary type. The auxiliary type is thought of as a second dominant type in that a person uses it more than any other type except for the dominant type. As with the dominant type, the auxiliary type is readily used and a person will unconsciously shift back and forth between the two. In the following possible combinations: ST, SF, NF, and NT (Fig. 1), the first letter of each type represents the dominant type and the second letter represents the auxiliary type.

Upon the completion and assessment of all the participants' MBTI profiles, the participants were paired according to three different categories: alike, opposite, and diverse.

'Diverse'group: these are pairs of participants who are alike in EITHER their dominant OR auxiliary preferences but not both

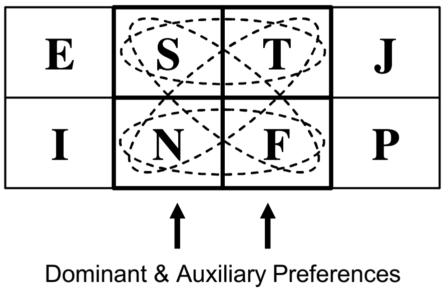

Fig. 1. MBTI dominant and auxiliary preferences.

(ST-SF, NT-NF, ST-NT, and SF-NF). Labeled as [divrs] in this experiment.

'Alike' group: these are pairs of participants who are alike in BOTH their dominant AND auxiliary preferences (ST-ST, NF-NF, NT-NT, and SF-SF). Labeled as [alike] in this experiment.

'Opposite' group: these are pairs of participants who are completely OPPOSITE in BOTH their dominant and auxiliary preferences (ST-NF and NT-SF). Labeled as [opp] in this experiment.

For assessing the level of communication skills, the instruments that had been used in some previous communication skill researches (Spitzberg, 2002; Spitzberg, 1997; Spitzberg and Cupach, 1984) were utilized in this experiment. Every participant was rated using the Conversational Skills Rating Scale (CSRS) form (Spitzberg, 1997) which served to evaluate the partner's communication skills. In labeling a person's communication skill level, the participants were divided into percentiles, 40% and 60%, in order to create a low level (<40%), a mid level, and a high level (>60%) of communication skill groups.

During the pairing process, a significant effort was made to bar influential factors such as grade point average and prior programming experience. All participants were found to have no prior programming experience, hence no special arrangement was made for that factor. For the grade point average factor, participants with similar grades were paired. For example, a student with an "A" grade was paired with a student with either an "A" or "B+" grade.

The four programming problems have been developed and validated by three independent professional programmers for parity, difficulty, and fitness to the experimental scope. Before devising the problem set, they reviewed the participants' profiles and pro-

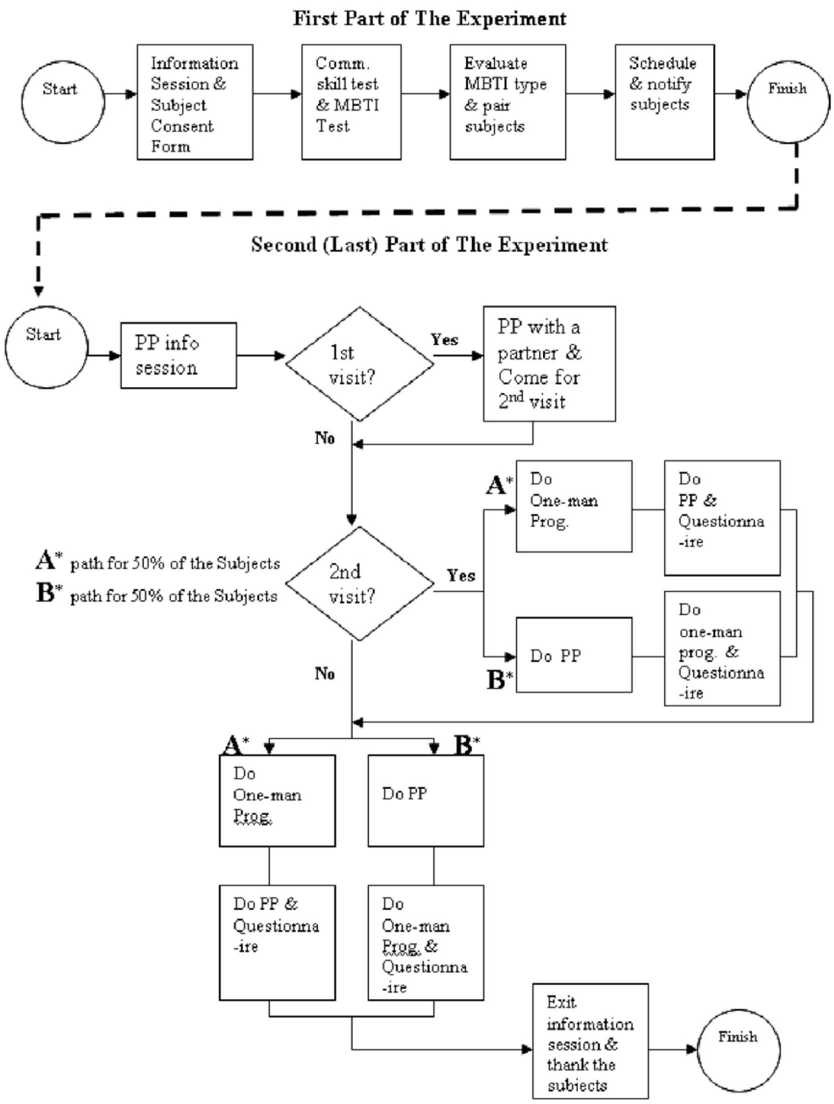

Fig. 2. Experiment flow.

gramming backgrounds and were fully informed about the experiment's scope and flow.

As Fig. 2 illustrates, the experiment had the participant pairs go through a total of three visits, barring any unforeseeable emergencies. In the real world, a software development effort goes on for months or sometimes for years. In this experiment context, the participants participated in several programming sessions in an effort to more closely resemble a regular software project development effort. Once the pairs were identified, the three scheduled sessions were arranged, as much as possible in consecutive weeks.

Each of the first two visits consisted of a 90-min programming session, where the participant programmed alone for 45 min and programmed in a PP environment for 45 min. The order of individual programming and PP was alternated in the two visits. Also, 50% of the sample pool started by programming alone and the other 50% started with PP. In 'Programming Alone' mode, each participant was given one of four problems for programming. The participants were instructed of the 45-min time limit and to turn in hard copies of their work at the end. In 'Pair Programming' mode, the pair of assigned partners was given another problem out of the four possible problems that neither partner had done. Again, the participants were told that 45 min were assigned and to turn in hard copies of their work. At the end of each visit, a post-session questionnaire was given to the participants. Additionally, the PP sessions were video recorded for analysis.

In the third visit, the participants were provided with the information that had been hidden from them. Lastly, they were thanked for their participation.

4. Analyzing the result

The questionnaires were collected and analyzed for factor analysis and Cronbach reliability measurement (Straub, 1989; Rosenthal & Rosnow, 1991; Cohen and Swerdlik, 2002). Through an extensive literature review, a handful of research with a focus on the psychosocial aspect of PP (Williams, 2000) was found. Consequently, the questionnaire items had to be originally designed for this experiment., The constructs of interest in this experiment are PP satisfaction, PP compatibility, PP communication, and PP confidence. In scoring the answers the pro factors (satisfaction, compatibility, communication and confidence) are weighed from seven ("strongly disagree") to one ("strongly disagree"). For the anti-factors (satisfaction, compatibility, communication and confidence), involving questions 4, 6, 7, 8, 13, 15, 16, 20, 21, 25, 26, 27, 28, 30, 31, 34, and 35, the weighing is reversed. Lastly, four short answer items were developed.

The 37 Likert scale question items were checked for validity and reliability by performing factor analysis and Cronbach reliability measurement. The questionnaires from all participants were collected and the results were manually recorded into a Microsoft Excel worksheet. The factor analysis was used by means of principal components analysis with varimax rotation. The parameters are (1) eigenvalues greater than one and (2) maximum iteration for convergence set to be 25. Typically, factor analysis is repeated until all item values are acceptable. For example, after a factor analysis, any item's value with less than 0.500 is discarded (Straub, 1989), as is any construct with less than a minimum of three items. Only after all item values are 0.500 or higher and all constructs possess a minimum of three questions is the factor analysis complete. Cronbach reliability measurement indicates how well a set of items measures a single latent construct. Typically, 0.700 is viewed as the acceptable minimal value and the higher the value, the more reliable the set (Cohen and Swerdlik, 2002). The following table is the matrix after the first round of factor analysis.

In validating the constructs, the first exploratory factor analysis is performed on all 37 items and determines which items load together for how many factors (constructs). In Table 1, ten factors are shown. Of those, items that carry a value of less than 0.500 and situations where only two items load together are discarded.

- Items 28, 16, 21, 25 – These items are discarded because they have no value with 0.500 or higher.
- Items 24, 22, 37, 36, 27, 35, 5 – These items are discarded because they don't satisfy the condition of having a minimum of three items that are required to form a set (construct).

This leaves six factors (1, 2, 3, 4, 5, 8). A second round of factor analysis is performed, shown in Table 2.

From the second round of factor analysis, items 15, 18, 17, 23 are discarded for the following reasons:

- Items 15, 23 – These items are discarded because they have no value with 0.500 or higher.
- Items 18, 17 – These items are discarded because they don't satisfy the condition of having a minimum of three items that are required to form a set (construct).

This leaves five factors (1, 2, 3, 4, 5). A third round of factor analysis is performed, shown in Table 3.

In this third round of factor analysis, all items meet the requirements. As the matrix table illustrates, items 32, 34, 33, 30, 31, 29, 19 load to factor 1, items 7, 6, 8, 4, 12 load to factor 2, items 10, 9, 11, 14 load to factor 3, items 3, 1, 2 load to factor 4, and items 13, 26, 20 to factor 5. From the 5 factors or 5 constructs, the appropriate construct names are PP communication for factor 1, PP satisfaction for factor 2, PP confidence for factor 3, programming alone confidence for factor 4, and PP compatibility for factor 5. The one-man programming confidence construct is dropped before further analysis. All items conform to their construct name. But item 19 does not quite correlate with its construct name, PP communication. Item 19 in the PP communication construct is kept as its factor analysis and reliability values belong to the construct. The Crobach's Alpha measurement for each set displays above 0.700, shown in Table 4.

As an example, the Likert scale on each question was designed as follows:

In PP, my partner insisted in doing thing his way and/or did not collaborate	SD* 1–2–3–4–5–6–7	SA**

*SA, strongly agree; **SD, strongly disagree.

From a 1–7 scale, the participants were asked to clearly circle a number that best represents his or her perception on the corresponding question.

Before analysis, the questionnaire results were checked for normal data distribution. The result as Table 5 shows does not exhibit normal data distribution. Only PP confidence is showing normal distribution ($p = .076$). However, for consistency, all constructs were analyzed by a non-parametric test, Kruskal–Wallis test.

Table 6 shows that there are no significant differences among the means of the three different MBTI personality combination groups – [divrs], [opp], and [alike] – in PP communication ($p = .312$), PP satisfaction ($p = .109$), PP confidence ($p = .136$), and PP compatibility ($p = .305$). This result exhibits that the different MBTI personality combination pair groups make no significant difference on pair communication, satisfaction, confidence, and compatibility. Therefore the pair group contrasts are not performed.

Table 1
Rotated component matrix

Survey questions	Factor loadings									
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9	Factor 10
Item 32	.804									
Item 34	.743									
Item 30	.736									
Item 31	.732									
Item 33	.723									
Item 29	.718									
Item 19	.569									
Item 28*										
Item 6		.884								
Item 7		.863								
Item 8		.788								
Item 4		.746								
Item 12		.567								
Item 14		.512								
Item 10			.820							
Item 11			.790							
Item 9			.710							
Item 15*										
Item 13				.767						
Item 26				.575						
Item 20				.570						
Item 16*										
Item 21*										
Item 3					.833					
Item 1					.767					
Item 2					.746					
tern 24						.751				
Item 22						.627				
Item 37							.817			
Item 36							.785			
Item 18								.618		
Item 23								.538		
Item 17								.522		
Item 25*										
Item 27									.803	
Item 35										.742
Item 5										.546

Extraction method: Principal component analysis.
Rotation method: Varimax with Kaiser normalization.
A rotation converged in 24 iterations.
* <.500.

Regarding the pair groups in terms of their level of communication skills, a test was done to detect any significant difference among the different levels of communication skill combination pair groups [HH], [HL], and [LL]. In Table 6, PP satisfaction ($p = .217$), PP confidence ($p = .671$), and PP compatibility ($p = .643$) reveal that the means of the three different levels of communication skill combination pair groups are not significantly different from each other. However, PP communication shows a significant difference ($p < .05$). To determine which combination pair group is significantly different from other groups, the Mann–Whitney test (Table 7) was performed.

Next, the gender factor was tested to see if any different gender combination pair groups reveal any significant differences. In Table 6, PP communication ($p < .05$), PP satisfaction ($p < .1$), and PP compatibility ($p < .001$) show significant differences among the three pair groups [MM], [MF], and [FF]. This leads to further analysis on how the three pair groups are significantly different from each other in PP communication, PP satisfaction, PP confidence, and PP compatibility.

As Table 7 exhibits, [HH] has a significantly higher mean than both [HL] and [LL]. Evaluated against [HL], [HH] shows $p = .027$ and against [LL], [HH] shows $p = .008$. However, in [HL] versus [LL], no significant difference is found as $p = .115$. This finding implies that only the homogenous [HH] group generates a high level of communication among pairs.

As the first construct, a test was done to see if there was any significant difference among the three pair groups in PP communication. In Table 8, [MM] shows a significantly higher mean ($p < .01$) than [MF], [FF] shows a significantly higher mean ($p < .01$) than [MF], and there is no significance in means ($p = .144$) between [MM] and [FF]. This indicates that two individuals of the same gender tend to perceive that pair communication is significantly different or better than when the two individuals are of different genders.

In regards to the PP satisfaction construct in Table 8, [MM] exhibits a significantly higher mean ($p < .05$) than [MF], and [FF] shows a significantly higher mean ($p < .1$) than [MF]. However, [FF] does not show a significant difference from [MM]. Similar to PP communication, PP satisfaction also reveals that two individuals of the same gender tend to perceive a higher level of pair satisfaction than the two individuals of different genders.

For the PP compatibility construct, shown in Table 8, [MM] exhibits a significantly higher mean ($p < .001$) than [MF], and [FF] shows a significantly higher mean ($p < .05$) than [MF]. Again, the pair groups of the same gender perceive a higher level of compat-

Table 2
Rotated component matrix II

Survey questions	Factor loadings					
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Item 32	.794					
tern 34	.767					
Item 33	.748					
Item 30	.734					
Item 31	.728					
Item 29	.685					
Item 19	.523					
Item 7		.907				
Item 6		.899				
Item 8		.793				
tern 4		.718				
Item 12		.566				
Item 14		.523				
Item 10			.838			
Item 11			.777			
Item 9			.711			
Item 15*						
Item 3				.855		
Item 1				.811		
tern 2				.731		
Item 13					.774	
Item 26					.692	
Item 20					.643	
Item 18						.739
Item 17						.634
Item 23*						

Extraction method: Principal component analysis.
Rotation method: Varimax with Kaiser normalization.
A Rotation converged in 8 iterations.
* <0.500.

ibility than the pair groups of different genders. [FF] and [MM] do not show significant differences ($p = .210$) to each other.

5. Discussion

Based on this experiment's results, the personality factor has no impact on the pair's satisfaction, communication, confidence, and compatibility. The levels of pair satisfaction, communication, confidence, and compatibility are not significantly different among similar MBTI personality type pair groups, opposite MBTI personality type pair groups, and diverse MBTI personality type pair groups. This is in contrast to what the survey revealed, where most of the results point to personality as the most influential factor to successful pair programming. This incongruity can be explained by supposing that the respondents meant that a person's modesty, manners, and respectability, rather than a certain personality type, is most conducive to a successful working relationship.

As expected, the [HH] pairs exhibited significantly higher levels of communication than [HL] and [LL], but did not show the same differences with regards to satisfaction, confidence and compatibility. This result contradicts the common belief that clear and well-delivered messages are likely to increase a person's satisfaction, confidence and compatibility. The result of this experiment suggests that regardless of how clear the communication is, it may bear no impact on other intangibles such as confidence, compatibility and satisfaction.

For the gender factor, as shown through many previous gender variable research studies, there is a significant difference in the areas of communication, satisfaction and compatibility. As expected, the same gender pair groups, [MM] and [FF], showed higher levels of communication, satisfaction and compatibility than [MF]. However, it has been found that the gender factor has absolutely no impact on pair programming productivity (Choi, Deek, & Im, 2008). This is relevant in the two dimensions of group collaboration, the business side and the personal side. In official business conduct, a factor such as gender has no bearing on the business result or outcome. However, on a personal level, a partner of the same gender does have some degree of influence on a person's psychosocial comfort zone. Based on this evidence, it would be safe to say that we can expect a much higher level of group cohesion if the group is composed of members of the same gender, and this fact would be physically more explicit with [FF] pair group.

An intriguing discovery is that the level of confidence on the finished work is not on par with the three other constructs. The same-gender pair group enjoys high levels of communication, satisfaction, and compatibility, but not confidence, on completed work. This underpins the aforementioned statement that the gender factor has no weight within the business dimension of group collaboration. Although the group members are content, they may have some doubts and different expectations about the group-assigned task.

Decoding the video record of PP sessions displayed more evidence of same gender affinity. The facial expressions, proximity of the partners, body gestures, and other noticeable physical behaviors were reviewed. Because the video recording portion was a supplement to the experiment data set, a general review and assessment is made instead of a full analysis.

There were some noticeable behaviors. Female–female pairs showed more pair closeness than other pairs. They shared more smiles and closer sitting arrangements at the keyboard. Also, they were more animated with their hand gestures and facial expressions. The male–male pairs did not exhibit behaviors similar to the female–female pairs. The male–male pairs' behaviors were mostly similar to the male–female pairs, with very limited facial expressions, no hand gestures, and sitting arrangements

Table 3
Rotated component matrix III

Survey questions		Factor loadings				
		PP corn in.	PP satisf.	PP conf.	One-Man Prog.	PP cornpat
32	In PP, my partner's active communication to me allowed me to be more active in expressing my views as well	.790				
34	In PP, my partner's hand gestures, eye gaze, body positions and other communication cues were NOT used adequately and poorly managed in our communication	.765				
33	In PP, my partner's voice tone was loud and clear which helped our communication	.742				
30	In PP, my partner did not express nor communicated much (too quiet) which made PP very difficult	.735				
31	In PP, my partner's message delivery was unclear which made PP very difficult	.732				
29	In PP, my partner described his or her point very well and I was able to fully understand	.692				
19	In PP, I really enjoyed my partner's partnership (willing to work as a team, open mind, etc)	.511				
7	I feel that I accomplish more when programming alone		.904			
6	I program best when I'm left alone		.900			
8	Compared to PP, I was able to produce a better solution when I programmed alone		.792			
4	Compared to PP, I enjoyed the one-man programming session more		.719			
12	I enjoyed the PP session more than programming alone		.574			
10	I believe that our code is readable that others can follow and understand it with no problem			.856		
9	I am confident that our work (output) is done correctly			.791		
11	I believe that our code is efficiently written (with less number of lines)			.783		
14	In PP, I believe that I had an easier time reaching the solutions			.557		
3	My code is efficiently written (with less number of lines).				.865	
1	I am confident that work (answer) is done correctly				.790	
2	My code is readable that others can follow and understand with no problem				.761	
13	In PP, My partner and I disagreed frequently in re aching a solution					.785
26	In PP, there were times when I was withdrawn (or maybe upset) because of the disagreements from the partner					.689
20	In PP, my partner insisted in doing things his way and/or did not collaborate					.649

Extraction method: Principal component analysis.
Rotation Method: Varimax with Kaiser Normalization.
A Rotation converged in 8 iterations.

Table 4
Cronbach's reliability measurement

Constructs	Reliability (Alpha)
PP comtn.	0.8794
PP satisf.	0.8909
PP conf	0.8314
PP compat.	0.7084

Table 5
Tests of normality of constructs

Construct	Kolmogorov–Srnirnov (K–S) test		
	Statistic	dE	Sig.
PP comm.	.112	210	.000
PP SatisE	.082	210	.002
PP ConE	.059	210	.076
PP Compat.	.121	210	.000

apart from each other at the keyboard. There were also some pairs that had a participant who was more dominant than the other. The dominant participant would control the "driver" seat throughout the PP session by either directly driving the keyboard or overwhelming the partner with his code review advices and checks.

Table 7
Test statistics for communication skill level – PP communication

	[HH] vs [HL]	[HH] vs. [LL]	[HL] vs. [LL]
Mann–Whitney U	763.0	175.000	824.5
Wilcoxon W	4166.0	47S.000	1124.5
Z	−1.93	−2 420	−1.204
Asymp. Sig. (2-tailed)	.054	.016	.229
Asymp. Sig. (1-tailed)	.027	.008	.115

Table 6
Test Statistics

	MBTI Personality				Comm. Skill Level				Gender			
	C	S	F	P	C	S	F	P	C	S	F	P
Chi-Square	2.33	4.42	3.99	2.37	6.436	3.053	0.798	0.884	12.105	5.745	0.984	19.209
Df	2	2	2	2	2	2	2	2	2	2	2	2
Asymp.Sig.	0.312	0.109	0.136	0.305	0.040	0.217	0.671	0.643	0.002	0.057	0.612	0.000

C, PP Comm.; S, PP Satisf.; F, PP Conf.; P, PP Compat.

Table 8
Test statistics for gender

	PP Communication			PP Satisfaction			PP Compatibility		
	1	2	3	1	2	3	1	2	3
Mann–Whitney U	3596.5	444	958.5	3819.5	564.5	1039	3255	482	1136
Wilcoxon W	6836.5	3684	8098.5	6900.5	3645.5	8299	6576	3803	8639
Z	−2.921	−2.808	−1.063	−2.184	−1.604	−0.619	−4.123	−2.797	−0.495
Asymp. Sig. (2-tailed)	0.003	0.005	0.288	0.029	0.109	0.536	0	0.005	0.621
Asymp. Sig. (1-tailed)	0.002	0.003	0.144	0.015	0.054	0.268	0	0.003	0.210

1 = [MM] vs. [MF], 2 = [MF] vs [FF], 3 = [MM] vs [FF].

6. Limitations

Although two individuals cognitively collaborate in writing computer codes together, the personality factor was not found to be an impact in the environment of this experiment's two 45 min pair programming sessions. However, this result can be interpreted in two different ways. One way is to take the result at face value and the other is to validate the result through a modified second-run with more visits consisting of much longer pair programming sessions. Another shortcoming is the completeness of the questionnaire. Because it had to be created, the general consensus is that the number of questions is too small, just three in some constructs, to fully extract thorough responses. Moreover, the use of non-parametric tests, due to data not being normally distributed, limited the analysis. The student participant pool may also lessen this experiment's findings. Using university students with no or limited programming experience as participants and conducting just three visits are not the ideal attributes for experiments such as this, and it is also a complete injustice to the programming profession. Ideally, using experienced professional programmers with a few years of pair programming or team programming experience and running the experiment for a period of several months would have served better for this experiment. The pool of professional programmers may have yielded more insights and underlying subtleties on the psychosocial factors and, more importantly, their magnitudes. Lastly, using a personality assessment tool besides the MBTI is another possibility.

7. Conclusion

Related to this study, an earlier related field survey study was undertaken (Choi, 2007). A group of professional programmers and programming shop management was asked: "What is the most influencing factor(s) to the pair programming?" The most popular answer was personality. However, we found that in this experimental setting, the personality was not significantly related to pair programming satisfaction, communication, confidence, and compatibility.

Similar to the findings of many previous gender-focused studies, this experiment's results affirm the belief that the same gender brings a higher level of group cohesiveness. There is a strong affinity among same-gender pairs which significantly impacts the group's perceptions of communication, satisfaction, and compatibility. This implies that a team's cohesion may be elevated if the same gender is used. Nonetheless, the same-gender team did not show a similar level of perception of confidence on the finished product. From management's perspective, if confidence about the finished product is not a concern in a given situation, then allowing the same gender to make up the team would result in a higher level of satisfaction, communication, and compatibility.

References

Amason, A. C., & Sapienza, H. J. (1997). The effects of top management team size and interaction norms on cognitive and affective conflict. *Journal of Management, 23*(4), 495–516.

Bayne, R. (1995). *The Myers–Briggs type indicator: A critical review and practical guide* (1st ed.). London, United Kingdom: Nelson Thornes Ltd.

Beck, K. (2000). *Extreme programming explained: Embrace change*. Reading, MA: Addison-Wesley.

Bradley, J. H., & Hebert, F. J. (1997). The effect of personality type on team performance. *Journal of Management Development, 16*(5), 337–353.

Bishop-Clark, C. (1995). Cognitive style, personality, and computer programming. *Computers in Human Behavior, 11*(2), 241–260.

Cheng, M. M., Luckett, P. F., & Schulz, A. (2003). The effects of cognitive style diversity on decision-making dyads: An empirical analysis in the context of a complex task. *Behavioral research in accounting, 15*(1), 39–62.

Choi, K. (2007). Team programming influencing factors: A field survey. *Journal of International Technology and Management*, *XVIII* (N3–4), 1–13.

Choi, K., Deek, F. P., & Im, I. (2008). Exploring the underlying aspects of pair programming: The impact of personality. *Journal of Information and Software Technology*, *50*(11), 1114–1126.

Cohen, R. J., & Swerdlik, M. E. (2002). *Psychological testing and assessment: An introduction to tests and measurement* (5th ed.). Boston: McGraw-Hill.

Da Cunha, A. D. & Greathead, D. (2004). Code review and personality: Is performance linked to MBTI type? http://eprints.ncl.ac.uk/file_store/trs/837.pdf.

Gallavan, N. P., & Juliano, C. M. (2007). Collaborating to create future societies with young learners. *Social Studies and the Young Learner, 19*(4), 13–16.

Katira, N., Williams, L., Wiebe, E., Miller, C., Balik, S., & Gehringer, E. (2004). On understanding compatibility of student pair programmers. In *Proceedings of the 35th SIGCSE technical symposium on Computer science education SIGCSE '04. 2004*, (pp. 7–11) Norfolk, Virginia.

Keirsey, D. (1998). *Please understand me II: Temperament, character, intelligence* (1 st ed.). Del Mar, CA: Prometheus Nemesis.

Mayer, R. E. (1981). The psychology of how novices learn computer programming. *ACM Computing Surveys, 13*(1), 121–141.

Maznevski, M. L, & Distefano, J. J. (2000). Global leaders are team players: Developing global leaders through membership on global teams, *Human Resource Management, 39*(2–3). 195–208.

Myers, I. B., & Myers, P. B. (1995). *Gifts differing: Understanding personality type*. Palo Alto, CA: Davies-Black Pub.

Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd ed.). McGraw-Hill.

Savicki, V., Kelley, M., & Lingenfelter, D. (1996). Gender, group composition, and task type in small task groups using computer-mediated communication. *Computers in Human Behavior, 12*(4), 549–565.

Schellens, T., & Valcke, M. (2005). Collaborative learning in asynchronous discussion groups: What about the impact on cognitive processing? *Computers in Human Behavior, 21*(6), 957–975.

Spitzberg, B. H. (1997). *A user's guide to relational competence and related measures*, unpublished.

Spitzberg, B. H. (2002). *Methods of interpersonal skill assessment*, *The handbook of communication and social interaction skills* (pp. 93–134). Mahwah, NJ: Erlbaum.

Spitzberg, B. H., & Cupach, W. R. (1984). *Interpersonal communication competence*. Beverly Hills, CA: Sage.

Stewart, L. P., Cooper, P. J., Stewart, A. D., & Friedley, S. A. (1996). *Communication and gender* (3rd ed). Scottsdale, AZ: Gorsuch Scarisbrick. 1996.

Stotts, D., Williams, L., Nagappan, N., Baheti, P., Jen, D., & Jackson, A. (2003). Virtual teaming: Experiments and experiences with distributed pair programming. *Extreme Programming/Agile Universe 2003* (pp.129–141). New Orleans, LA.

Straub, D. W. (1989). Validating Instruments in MIS Research, *MIS Quarterly*, *13*(2), 147–169.

Tannen, D. (1991). *You just don't understand: Women and men in conversation*. New York: Ballantine Books.

Thomas, L., Ratcliffe, M., & Robertson, A. (2003). Code warriors and code-a-phobes: A study in attitude and pair programming. In *Proceedings of the 34th SIGCSE technical symposium on Computer science education* (pp. 363–367). Reno, Nevada, USA.

Thompson, L., & Hastie, R. (1990). Social perception in negotiation. *Organizational Behavior and Human Decision Processes, 47*(1), 98–123.

Tor, Busch, 1996. Gender, group composition, cooperation, and self-efficacy in computer studies, *Journal of Educational Computing Research*, *15*(2), 125–135.

Weinberg, G. M. (1998). *The Psychology of Computer Programming* (Silver anniversary ed). New York: Dorset House Pub.

Williams, L. (2000). *Collaborative Software Process*, PhD Thesis. Available from: <http://www.cs.utah.edul-lwilliamlPaperslddissertation.pdf>. Retrieved on 25 November 2003.

Williams, L., & Kessler, R. R. (2000). All I really need to know about pair programming I learned in kindergarten. *Communications of the ACM, 43*(5), 108–114. May 2000.

Williams, L., & Kessler, R. R. (2002). *Pair programming illuminated*. Reading, MA: Addison-Wesley.

Winter, Janet; Neal, Joan C.; Waner, & Karen K.(2005). Student teams learning to cope with conflict, *Delta Pi Epsilon Journal, 47*(2) (pp. 67–74).

Yourdon, E. (1997). *Death march*. Upper Saddle River: N.J. Prentice-Hall.