

The Impact of an Agile Methodology on the Well being of development Teams

Sharifah Syed-Abdullah, Mike Holcombe, Marian Gheorge,

Department of Computer Science, University of Sheffield, UK

s.abdullah@dcs.shef.ac.uk

Abstract

This paper describes an empirical study, which addresses the aspect of well being amongst members of the software development teams. The question of interest is whether an agile methodology has any distinct effect on the well being of the software developers. Both quantitative and qualitative methods were utilised, including the participative observation, focus group interviews, close-ended questionnaires and simple statistical tests such as Spearman Correlation and Mann-Whitney test. Initial results showed that an agile methodology (XP) has a positive effect on the level of enthusiasm of the software developers in the most dynamic project. To understand why XP can increase enthusiasm, results are interpreted with references to cognitive, affective and managerial properties of the practices studied. This result needs further investigation on the individual effects of each practice on the wellbeing and attitudes of Software Engineering (SE) teams.

1 Introduction

Adherent of agile methodologies claim that they are more adaptive than predictive and more people-oriented than process-oriented. Adaptive approaches said to be better when the requirements are uncertain or volatile. If the requirement is not stable perhaps because of business dynamics and rapidly evolving market pressure, it is difficult to develop stable design and followed a planned process as being practiced in traditional methodologies. As a reaction to these problems, a new group of methodologies known as **agile methodologies** evolved. Agile methodologies welcome change and unpredictability in software projects. Among the agile methodologies, Extreme Programming (XP)(Beck 2000a) is the one that has received the most attention. This is because of the ability of XP leaders to refine their practices on numerous projects during the early 90's, extending their ideas of an approach that are both adaptive and people-oriented. XP begins with 4 values; *Communication, Simplicity, Feedback* and *Courage*. It then builds up to a dozen practices, which XP projects should follow. The XP

practices were developed and tested around the humanistic phrase that “*software is built by human beings. In the end keep the human being focused, happy, and motivated and they will deliver*” – Kent Beck (Beck et al. 2001)

The study described in this paper addressed the humanistic aspects of the Extreme Programming with a particular emphasis on the work related wellbeing of the developers in three different levels of project environments and how this was influenced by the intrinsic relationship of the 12 practices in Extreme Programming (XP). In this paper, we examined 17 software development teams, the methodologies used, the dynamism of the projects being developed and the wellbeing variables measured against the teams. The empirical study described in this paper was carried in the “Sheffield Software Engineering Observatory (Holcombe et al. 2003a) and followed as closely as possible the guidelines for empirical research in software engineering (Kitchenham et al. 2002).

2 Related Work

Work related well being

Research has suggested that four factors have a significant effect on the well being of an employee and these are job design, performance monitoring, human resource practices and team leader’s support (Frenkel et al. 1998; Knight et al. 1998). Studies on job design have demonstrated that control, variety and the demands placed on the employees are important predictors of well being. With regards to job control, a study has shown that high job control is positively associated with well being (Holman 2002) and job satisfaction (Batt et al. 1995). Findings by Holman showed that having high control over work methods and procedures, a low level of monitoring and a supportive manager, would appear to have the most significant effects on the employee’s wellbeing. This finding is in line with other research, which found that high level of monitoring has a negative association with wellbeing (Chalykoff et al. 1989).

XP methodology

XP was created in response to problem domains whose requirements change and to address the problem of project risk. The humanistic aspect of the communication and the simplicity

aspect of the methodology, promote good teamwork because it is an important ingredient towards developing quality software. A stable teamwork facilitated continuous testing, which enhanced courage because members were confident of producing better and well-tested software. The satisfaction of producing quality software, was very important because it boosted the confidence of the team to produce a more challenging software.

Discovery method

The Discovery Method (Simons 1998) is a methodology developed by a lecturer in the University of Sheffield. It provided a completely guided approach to modern object-oriented and component-based software development. The method covers the part of the software lifecycle ranging from initial requirements elicitation, through task, object and subsystem identification, to detailed coding specifications. A key feature is the exploitation of discovery procedures, self-reinforcing analytical techniques having a trigger, a feedback loop and a completion criterion. A linked sequence of discovery procedures forms a discovery path, a development route leading from one modelling stage to the next. It is a transformational method, in which analysis models evolve gradually into system designs with maximum cohesion and minimal coupling, and are integrated with existing libraries and application frameworks. It is a selective method in which techniques are chosen for their single focus and fitness-of-purpose in context. It is a participatory method, which the client is continually involved. The Discovery Method uses an adapted subset of UML notations, whose semantics have been clarified such that models have a formal underpinning and may be related to each other by transformation. The iterative, incremental and parallel development strategy may be described within the OPEN Process Specification.(Simons 1998; Simons 2003).

3 Motivation and Research questions

The motivation for the experiment was to investigate if the use of an XP methodology helps to improve the work related well being of the SE teams. In doing so, it was interesting to understand how the different practices in this methodology contributed to the result of the experiment.

3.1 Research questions and Hypotheses

The first objective of this study was to assess empirically, whether there are any differences in the work related well being level between the software development teams using an agile methodology and the teams using a designed-based approach. The research question was therefore as follows:

Research question 1:

Which software development methodology have a greater impact on the work related well being of the SE teams?

The hypothesis was defined as:

H_A: The agile teams will experience a higher level of well being compared to the designed-based teams.

Since the work-related wellbeing was measured using a scale, which consisted of four constructs, that are anxiety-contentment and depression-enthusiasm, the above hypothesis was further divided into four hypotheses:

H₁: Developers using the XP methodology will experience a lower level of anxiety than those using the designed-based methodology.

H₂: Developers using the XP methodology will experience a higher level of contentment than those using the designed-based methodology.

H₃: Developers using the XP methodology will experience a lower level of depression than those using the designed-based methodology.

H₄: Developers using the XP methodology will experience a higher level of enthusiasm than those using the designed-based methodology.

Besides exploring the impact of the two types of methodologies on work related methodology, the experiment also serve to evaluate the relationship between the number of the practices in the XP methodology and the level of work related well being level of the development teams.

Research question 2:

Is there a relationship between the number of XP practices used and the level of the work related well being experienced by the SE teams?

The hypothesis was defined as:

H_B: The higher number of XP practices adopted are associated with a higher level of work related well being experience by the members.

The hypothesis was further divided into 4 separate hypotheses:

H₆: The higher the number of XP practices used, the lower the reported anxiety level will be

H₇: The higher the number of XP practices used, the higher the reported contentment level will be

H₈: The higher the number of XP practices used, the lower the reported depression level will be

H₉: The higher the number of XP practices used, the higher the reported enthusiasm level will be

4 The Experimental Design

4.1 Context of the Experiment

Subjects.

Subjects were the 2nd year undergraduate students from Computer Science and Software Engineering degrees, and the 3rd year students from Mathematics and Computer Science degree, who were taking the Software Hut module (Holcombe et al. 2003b). This involved students working on the real industrial projects with the external clients. Each client then worked with 5 to 6 distinct teams who designed and implemented a solution in competition with the other teams.

Group Assignment.

During the first week, the students were encouraged to choose their own software development team members. The formation of these groups in the Software Hut module was a combination of ‘natural’ groups and ‘constructed’ groups. Natural groups consisted of members who have been attending similar modules before attending the Software Hut module or members living in the same residential hall. The ‘constructed’ groups were teams, which have fewer team members and required the lecturers’ assistance to form a complete team. During this study, each team consisted of 4 to 5 members. With a total population of 75 students, there were 17 teams taking part in this study. The even-numbered teams were required to use the XP methodology and the odd numbered teams were required to use the Discovery method, a design-based methodology, which was studied during the previous semester. The Mann-Whitney statistical test conducted showed no significant difference in the average previous marks between the Discovery teams ($M = 59.11$, $SD = 7.441$) and the XP teams ($M = 58.75$, $SD = 6.042$). The freedom to define their own team members was often perceived, as a threat to validity of the experiment because there was a tendency for the higher grades students to group themselves together. In the longitudinal study conducted by the researcher, it was observed that given the chance to form their own team, the members were often grouped according to the personality traits of the members rather than the previous achievement. This observation was supported by the above statistical result. Therefore, the formation of group by the students themselves was not considered as a threat to validity in this experiment.

Project	Discovery team	XP team
Domestic Violence	1, 11, 17	6, 12, 14
Pharmaco	3, 5, 7	4, 8, 16
Control Engineering	9, 13, 15	2, 10

Table 1 The Distribution of team and project during the study.

During week 1, each client was required to give a brief introduction of their company and its requirement, to the students. After the briefing, the students were given a week, to compare skills among the team members with the skills needed to build a solution for each client. After the students made their evaluations based on their strengths and weakness, their choices were submitted to the lecturers. Final decision regarding clients-students teaming was made by the lecturers to ensure equal distribution (Table 1). The teaming was made as close as possible to the request to ensure that the students were satisfied with the decisions.

4.2 Setup of the Experiment

The experiment was based on three-time intervals; before treatment, during the treatment and after treatment.

Before Treatment

During the first week, the students were briefed, on the objective of the module and the competition between the teams developing the same project. The briefings does not emphasis on the different methodologies, but focused on the ability of the team to develop the best solution for the client. The students were informed, of the research work carried out by the researcher but the description of the research work was not revealed, therefore they were 'blinded' towards the purpose of this experiment. During the second week, all of the students were given a team management course to familiarise with managing a group project. The questionnaire was administered during this week because the assignment of the team was completed but the application of the methodology has not commenced. The questionnaire was administered to half of the team members because the other half of the team was involved in another experiment. The data collected represented the level of work related well being before the treatment.

During the Treatment

There were two hours time slot for this module. Throughout the next 4 weeks, one slot was allocated for lectures where the SE teams attended separate lecture and lab sessions. The XP teams were introduced to the methodology during the lectures and were coached on doing pair programming, test first coding and story cards during the lab sessions. When the XP teams

attended the lectures or lab sessions, the Discovery teams met the client. After the first hour, the Discovery teams attended the review sessions on Discovery method and the lecturer did monitoring on the application of this method. At this time, the XP teams were given the chance to attend the client meeting. In this way, each team had a chance of discussing their project for 20 minutes with the respective client.

After the 6th week, the lectures and lab sessions stopped. The teams were given the time to focus on developing the software. The two-hour sessions changed to meeting with the manager and meeting with the client. The rotation between XP teams meeting the client and the Discovery teams meeting the client was change to avoid the same team meeting a tired client, every week. During the client meeting, the XP teams were encouraged to show the developed part of the software as soon as possible.

After the Treatment

During the 11th week, the questionnaire was administered again, because this was the dateline for the teams to demonstrate and to deliver the developed software to the client. At this time, it was decided that the teams have sufficient time to experience using the respective methodology. The data collected after the methodology treatment was compared to the data collected before the treatment.

In total, there were three lecturers, who conducted the lectures (two lecturers for the XP sessions and one lecturer for the Discovery sessions), two coaches for the XP lab sessions, 1 trainer for the Discovery method, three managers for managing the SE teams (two of the lecturers acted as the managers) and 3 clients.

4.3 Variables

The experiment used five variables.

Methodology

There were two types of methodologies used. The first was the design-based methodology; the Discovery Method (Simons 1998) was selected because the developer of this methodology is a

lecturer in the same department. This facilitated the monitoring of the methodology during the experiment. The second methodology was an agile methodology; XP (Beck 2000b) was selected because it was the best known breed of the agile methodologies.

Project Rating.

The interview sessions, conducted among the development teams, an XP coach and a project manager, revealed that there were three levels of working environment for the projects. The client for the Domestic Violence project was able to define their requirements early in the project and therefore the project was rated as 1 to represent a stable environment. The client for the Pharmaco project had an initial basic requirement, which changed gradually as her understanding of the technology increased. The changes of the requirement warranted a rating of 2 to represent a dynamic environment. The Control Engineering project was rated as 3 to represent *unpredictable* environment due to the uncertainty about the preferences and technology required by the client. The lack of stability in the requirement specifications of this project seemed to serve as an impediment on the smooth progress of this project development.

Time interval

There were two time interval identified to collect the work related well being; before treatment and after treatment. The wellbeing questionnaire (Warr 1990) was administered during the second (before treatment) and the eleventh week (after treatment) of the semester. The face-to-face administration of the questionnaire was chosen, because the response rate was higher than the other methods. The first data collection was to get an early reading on the well being level among the teams. It was important to ensure that members of the same team answered within 1 week of each other. The sensitivity of the scale stated that if the difference were more than 1 week, the data would be invalidated.

The second reading was to measure the increase/decrease in well being level (if any). The difference between the readings was analysed, in order to study the effect of using XP or a design-based approach. The further apart the readings (beginning and ending of projects) were made; the better was the second reading. For the second reading, the questionnaire was

distributed to the students during the interview sessions conducted during the 11th week. The researcher was unable to get any response from one team during this week.

The work related well being

Job related anxiety, depression, contentment and enthusiasm were measured using the 12–items anxiety-contentment and depression-enthusiasm scales developed by Warr (Warr 1990). The scale was used to measure the extent to which the software team members were anxious or contented, depressed or enthusiastic about their project at the varying time intervals. Respondents were asked to think of the past few weeks and to indicate the extent to which they felt gloomy, calm, uneasy, enthusiastic, cheerful, worried, contented, tense, depressed, optimistic, relaxed and miserable. The validity and reliability of this scale has been demonstrated in other studies (Agho et al. 1992; Axtell et al. 2002; Sevastos et al. 1992). In this study, the wellbeing scale achieved an acceptable reliability level with a Cronbach’s alpha of 0.86. Cronbach alpha provides an assessment of the internal consistency of all the items in the scale used. Measurement was on a 5-point scale, and the respondents were asked to indicate the degree of agreement and disagreement with each item. Responses range from 1 “very slightly” or “not all” to 5 “very” or “extremely”. The response were totalled according to the 4 construct (APPENDIX B).The preliminary analysis was conducted to investigate whether there were any selection effects for these groups, that is those given to work with XP methodology may have been those with the most positive and least negative well being). T-tests revealed no differences between the groups on any of the dependent variables at week2

The number of the XP practices

The number of the XP practices used was measured through the interview sessions, analysis of the project documents and participative observations. The evidences were documented quantitatively using the table developed in an earlier study (Syed-Abdullah et al. 2003). Each practices was evaluated and assigned a percentage of the number of the XP practices used. The measurement was made against an XP activities table developed for this purpose (APPENDIX C).

A summary of the variables used in this experiment is defined in Table 2

Variable	Definition and source of measure used
Methodology	The 2 methodologies used for comparison <ol style="list-style-type: none"> 1. Discovery 2. XP
Project	The rating given to the project <ol style="list-style-type: none"> 1. stable project 2. dynamic project 3. unpredictable project
Time	The time intervals for collecting the data <ol style="list-style-type: none"> 1. before treatment (week 2) 2. after treatment (week 11)
Well being	The degree to which the members feel anxious, contented, depressed or enthusiasm about the project (Warr 1990)
Number of XP practices	The number of the XP practices applied when developing the project

Table 2 Definition of the variables

4.4 Other factors under consideration

The other factors that were considered during this experiment were size of the project, time spent by each team to develop the software and the quality of the developed software. The experiment was conducted together with another researcher who focused on the size and the time taken to complete the project. The following two section are the findings by the said researcher.

The size of the project

The assessment of the size of the project included the size of the code (K-characters), number of commentaries, the number of requirement and the number of test cases. “The use of the different programming languages allowed different possibilities to compact code and their expressiveness also affect the number of lines” [p. 89, (Francisco 2004)]. The study reported no large difference in the size of the project developed by the Discovery teams and the XP teams.

The time spent to develop the project

The time spent by the teams was registered and verified weekly. The study reported that the XP teams spend more time than the Discovery team to develop the respective project. Comparing the time taken between the two methodologies, the research reported that the XP teams spend more time on planning the testing, testing the code, coding and reviewing but less time in analysis of requirement and design activities. The time was higher in the XP teams because of the socializing and the communicating of the related activities. Nevertheless, the statistical tests conducted showed no significant differences between the two methodologies (Macias et al. 2003).

Software quality

The external quality was measured using 10 items (APPENDIX D) and the internal quality was measured using 6 items for the Discovery teams (APPENDIX E) and 7 items for the XP teams (APPENDIX F). The decision to focus on the external quality only was made, after considering the 'noise' factor associated with the internal quality. Analysis of the pattern in the internal quality revealed that by knowing the student's previous academic performance and the weekly performance, there was a tendency to look beyond the finished product to measure the teams' capabilities.

At the end of the term, the client chose the best software without being aware of the methodology used by the different teams. Before the final presentation, the clients were briefed on the structured marking scheme, which they were required to follow when awarding the marks to the teams. The clients were required to fill in the assessment document after every team had demonstrated and delivered their software. The average external quality for the two approaches were compared using the Mann-Whitney non parametric statistical test and the result showed *no significant difference* in the means score for the Discovery teams ($M=33.61$, $SD=6.314$) and the XP teams ($M=33.31$, $SD=5.398$).

5 Threat to validity

There were a number of threats to the validity of the experiment. First, the size of the sample was small therefore to achieve significant value was difficult. The decision to administer the questionnaire to only half of the team member was seen as a weakness to the study. Second, the means to measure the XP practices was collected qualitatively and the information gathered was then quantified. The process posed a threat to the validity because the data was not captured objectively. There is a need to identify experimental metrics to capture the number of practices used.

6 Analysis and Results

Independent sample t-test was used to compare the total mean score of the well being variable. When the data was split according to the projects, the non parametric tests were used because of the sample size was smaller. The Mann-Whiney test was used for measuring the difference between the methodologies and the Wilcoxon Signed Rank test was used for measuring the difference between the two intervals (week 2 and week 8).

6.1 Research Question 1: Impact of the Methodology on Work related Well being

Comparison according to the Methodology

Independent sample t-test showed no significant difference between the intervals for both methodologies; Discovery (N=20, M1= 46.60 SD = 5.46; M2 = 37.25, SD = 9.96) and XP (N=19, M1=46.74, SD = 5.98; M2 = 37.32, SD =7.79). The similarity in the mean score for wellbeing is illustrate in Figure 1

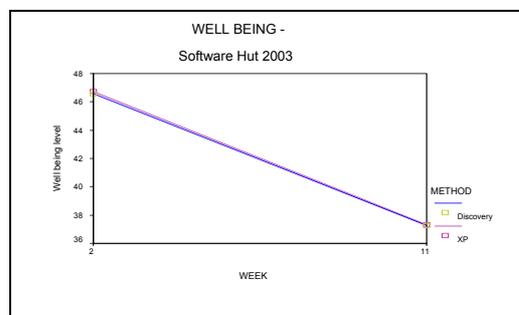


Figure 1 The well being measure between the two methodologies (Software Hut 2003)

To understand how the methodology affect the different feeling in the well being measure, the data was split into the four variables - anxiety, contentment, depression and enthusiasm (Figure 2).

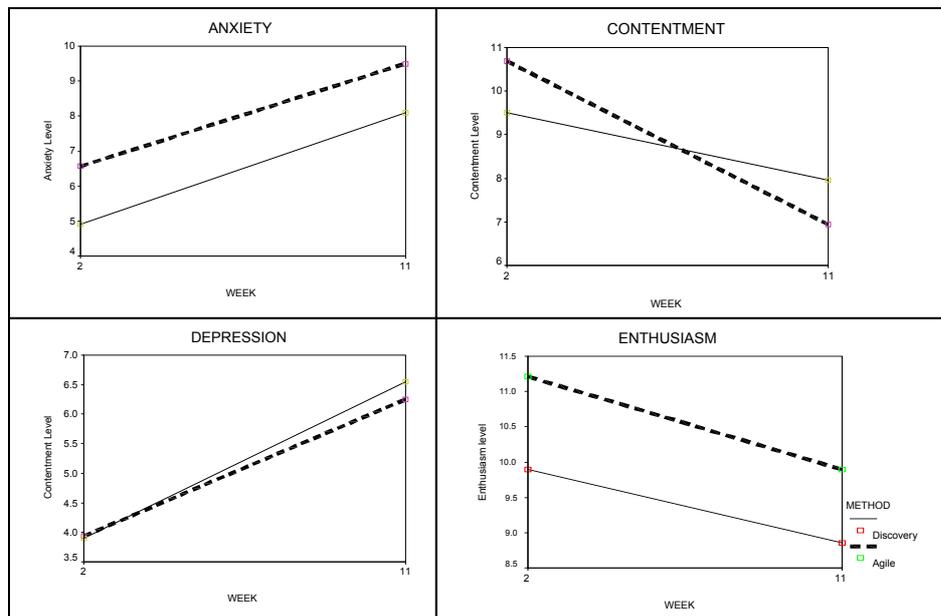


Figure 2 The difference in the feeling of anxiety, contentment, depression and enthusiasm experienced by the two teams (Software Hut 2003)

Anxiety level. There was a significant different between methodologies during the 2nd week and the 11th week. The Discovery teams ($M_1= 4.90$ SD = 1.62; $M_2 = 8.10$, SD = 3.24) experienced a lower anxiety level than the XP teams [($M_1=6.68$, SD = 2.69; $M_2 = 10.00$, SD =3.02); $t_1 = 2.495$, $p = 0.018$ * and $t_2 = 1.891$, $p = 0.066$].

Contentment level. There was no significant difference in the contentment level during the two intervals; Discovery teams ($M_1=9.50$ SD =3.035; $M_2 = 7.95$, SD = 2.911) and XP teams ($M_1=10.32$, SD = 2.358; $M_2 =6.74$, SD = 2.130).

Depression level. There was no significant difference in the depression level between Discovery teams ($M_1=3.90$ $SD = 2.29$; $M_2 = 6.55$, $SD = 4.045$) and XP teams ($M_1= 4.11$, $SD = 1.912$; $M_2 = 6.32$, $SD = 3.667$).

Enthusiasm level. There was a significant difference in the enthusiasm level during the 2nd week only, the Discovery teams ($M_1= 9.90$, $SD = 2.654$; $M_2 =8.85$, $SD = 2.159$) and the XP teams ($M_1=11.21$, $SD = 1.903$; $M_2 =9.89$, $SD =2.514$) ; $t_1 = 1.764$, $p = 0.086$].

Comparison According to the Project

The non parametric tests were used during this analysis; the Mann-Whiney test – for measuring the difference between the methodologies and Wilcoxon Signed Rank test – for measuring the difference between the two intervals (week 2 and week 11).

Anxiety level according to project

Figure 3 graphs the direction and magnitude of the effects of the methodologies on the anxiety level for the three types of project. Anxiety refers to the feeling of being tensed, worried and anxious in developing software. In the most dynamic project, it was observed that students using the XP methodology experienced lower level of anxiety at the end of the project.

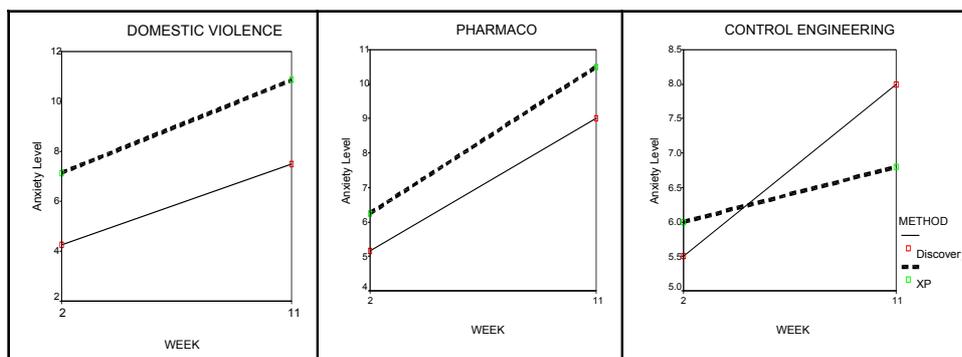


Figure 3 The difference in the level of anxiety according to project (Software Hut 2003)

For both the Domestic Violence and Pharmaco projects, anxiety level of the XP teams was higher than the Discovery teams but the statistical test conducted showed that only the teams developing the Domestic Violence project exhibited a significant difference [Discovery teams

($N=8$, $M_1=4.25$, $SD=1.03$; $M_2=7.50$, $SD=3.46$) ; XP teams ($N=7$, $M_1=7.14$, $SD=2.79$; $M_2=10.86$, $SD=1.68$) during the two intervals [$z_1=2.20$, $p=0.027$; $z_2=2.1$, $p=0.036$]. The possible reason for this result is due to the Discovery methodology treatment in the earlier course module.

However, the XP teams experienced a higher initial anxiety level for all of the projects. The teams had the disadvantage of having to learn the new methodology within a certain time frame in order to complete the project on time. The lack of knowledge had been perceived as a disadvantage in a competition that awarded the winning team with a computer laptop for each member. Previous research has shown that the perceived lack of knowledge has been associated with the increase in the anxiety level amongst the team members (Axtell et al. 2002; Clegg et al. 1997).

There was a significant positive relationship between methodology and anxiety for the Domestic Violence project, indicating that for a stable project a more structure methodology such as Discovery method was more suitable. This is to be expected as shown by previous research (Norman et al. 2003).

However in the unpredictable project, where the requirements were vague, the ability of the XP methodology to capture only partial requirements through the use of a simple story helped to prevent a drastic increase in the anxiety level. In addition, the advantage of coding the project in pairs and the benefit of frequently testing the code contributed to a less anxious environment. Nevertheless, the insignificant result failed to accept hypothesis H_1 . Eventhough, the finding cannot prove statistically that using the XP methodology lowered the anxiety level of the developers but the result provided an early evidence of the effect of different type of methodology in the different work environment.

Contentment level according to project

The contentment variables refers to the feeling of being contented, calm and relaxed when developing and completing the software project. Similar pattern was observed, in all of the three projects (Figure 4).

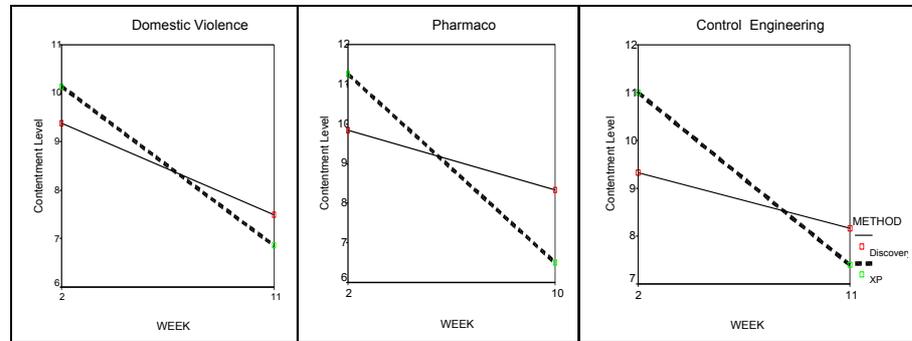


Figure 4 The difference in the level of contentment according to project (Software Hut 2003)

Both teams experienced a decrease in the level of contentment at the end of the project. Initially the XP members experienced a higher level of contentment but the rate of decreasing was much faster than the Discovery teams. The statistical test conducted showed that the XP teams developing the Pharmaco project ($M_1=11.25$, $SD=0.96$; $M_2=6.50$, $SD=2.08$, $z=1.84$, $p=0.06$) and the Control Engineering project ($M_1=11.00$, $SD=2.55$; $M_2=7.4$, $SD=2.07$, $z=2.041$, $p=0.041$) showed a significant decrease in the contentment level between the two intervals. Nevertheless, the comparison between the methodologies showed **no significant difference. Therefore the hypothesis (H_2) was rejected.** The finding failed to prove that using the XP methodology will cause the contentment level of the developers to be higher than the Discovery teams.

Depression level according to project

Depression refers to the feelings of gloom, being miserable and depressed towards the project. The XP teams developing the Domestic Violence and the Pharmaco projects experienced a higher depression level at the beginning of the project, nevertheless there was no significant increase between the two intervals. In the Control Engineering project where there was a higher level of uncertainty, the XP teams experienced lower depression level than their

counterpart. The statistical test conducted showed no significant difference between the methodologies and between the intervals.

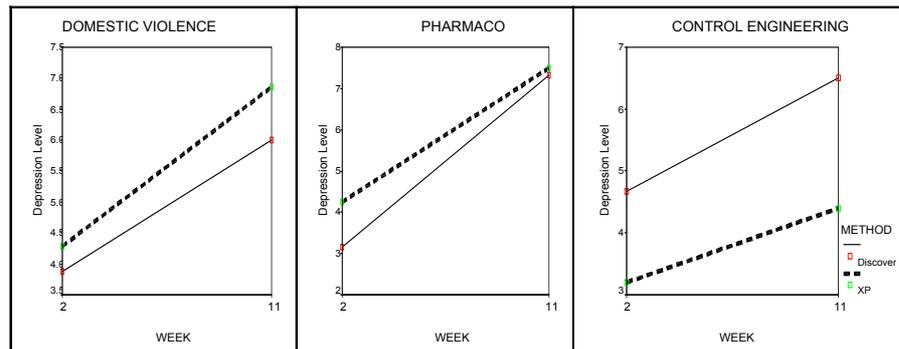


Figure 5 The difference in the level of depression according to project (Software Hut 2003)

The analysis of the depression variable provided the evidence that XP members experienced a lower depression level in the most dynamic project but the statistical test conducted showed **no significant different** between the methodologies; Discovery ($M_1=4.67$, $SD=3.141$; $M_2=6.50$, $SD=3.674$) and XP ($M_1=3.20$, $SD=0.447$; $M_2=4.4$, $SD=2.608$). **Therefore the third hypothesis (H_3) was not accepted.**

Enthusiasm level according to project

Enthusiasm is the measurement of feeling enthusiastic, optimistic and cheerful towards the project being developed.

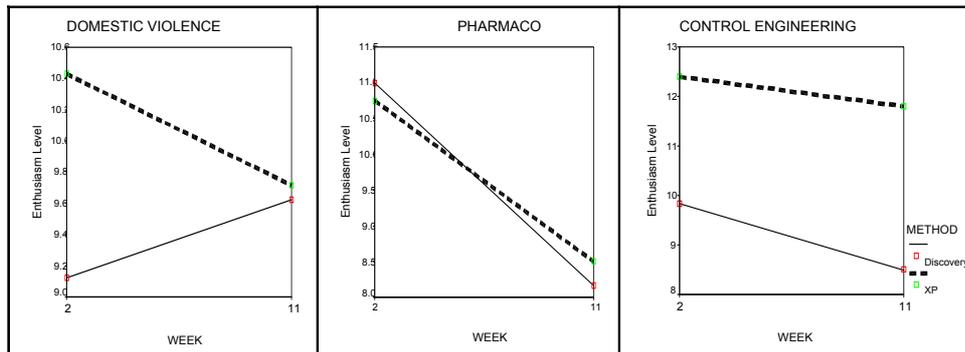


Figure 6 The difference in the level of enthusiasm according to project (Software Hut 2003)

The analysis of the enthusiasm variables revealed that XP members started with a higher level of enthusiasm for Domestic Violence and Control Engineering projects but the feeling started to decrease as the project progressed. At the end of the projects, the enthusiasm level was higher amongst the XP members than amongst the designed-based members. Interestingly, the enthusiasm level of the Discovery teams in the Domestic Violence project increased as the team progressed towards the end of the project. There **was a significant difference** in the enthusiasm level between the Discovery team ($M_2=8.50$, $SD=2.66$) and the XP team ($M_2=11.80$, $SD=2.17$, $z=1.93$, $p=0.054$). **The result supported the hypothesis (H₄) in the most dynamic work environment.**

6.2 Research Question 2: The relationship between the number of XP practices and the Work related Well being

To explore the relationship between the XP practices and the well being level of the development teams, four hypotheses were tested.

H₆: The higher the number of XP practices used, the lower the reported anxiety level will be

When the number of XP practices used was higher, the lower the anxiety level among the developers were reported (Spearman coefficient -0.579 , $p < 0.05$). This could be explained by a

whole myriad of factors ranging from the individual personality types of group members through to the specific human related aspects of the XP methodology.

H₇: The higher the number of XP practices used, the higher the reported contentment level will be

When the number of XP practices used is higher, higher contentment level among the developers were reported (Spearman coefficient 0.468, $p < 0.10$). This would suggest that a full adaptation of XP practices would yield a better outcome than an unholy amalgamation of methodologies.

H₈: The higher the number of XP practices used, the lower the reported depression level will be

When the number of XP practices used is higher, lower anxiety levels among the developers were reported (Spearman coefficient -0.293, $p < 0.05$). This backs up earlier research into happy SE teams (Melnik et al. 2002) and adds support to the claim that XP is a human centred approach to SE.

H₉: The higher the number of XP practices used, the higher the reported enthusiasm level will be

When the number of XP practices used is higher, higher enthusiasm levels among the developers were reported (Spearman coefficient 0.440, $p < 0.10$). This indicates that teams who gained a deep understanding of XP practices were able to utilise them in a more efficient manner than those who relied on ad-hoc methods.

The result of the hypotheses is summarized in the Table 3

Number	Hypothesis / Subhypotheses	Result/ Evidence
H_A	The agile teams will experience a higher level of well being compared to the designed-based teams	Accept in the unpredictable environment
H ₁	Developers using the XP methodology will experience a lower level of anxiety than those using the designed-based methodology.	Reject - but exist a strong support of a converse hypothesis. There was a weak support for the hypothesis in an unpredictable project
H ₂	Developers using the XP methodology will experience a higher level of contentment than those using the designed-based methodology.	Reject
H ₃	Developers using the XP methodology will experience a lower level of depression than those using the designed-based methodology.	Reject
H ₄	Developers using the XP methodology will experience a higher level of enthusiasm than those using the designed-based methodology.	Accept- Strong support in the unpredictable project. There was a weak support for a converse hypothesis in the stable project.
H_B	The higher number of XP practices adopted are associated with a higher level of work related well being experience by the members.	Accept
H ₆	The higher the number of XP practices used, the lower the reported anxiety level will be	Accept - Strong support
H ₇	The higher the number of XP practices used, the higher the reported contentment level will be	Accept -Strong support
H ₈	The higher the number of XP practices used, the lower the reported depression level will be	Accept -Strong support
H ₉	The higher the number of XP practices used, the higher the reported enthusiasm level will be	Accept -Strong support

Table 3 Summary of the hypotheses and result provided in this study

6.3 Discussion

Agile methodologies have been introduced primarily as an answer to the rapidly growing and volatile software industry. The study reveals that XP methodology has a positive impact on the wellbeing of the developers with the exception of contentment in the most volatile project. It has been shown that uncertainty and complexity increases the need for flexibility, adaptability and speed (Blomqvist et al. 2002). In this study, XP teams were shown ways to manage clients' uncertainty through the flexibility of story cards and the speed of releasing the functioning part of the system; while the project complexity was controlled through the use of simple design, refactoring, testing and pair programming.

In explaining the anxiety level, earlier exposure to design-based methodologies helped in reducing initial feelings of anxiety among design-based members. However in more dynamic projects, where the requirements were vague, the ability of the agile method to capture only partial requirements through the use of a simple story helped to prevent a drastic increase in anxiety levels. In addition, the advantage of coding the project in pairs and the benefit of frequently testing the code contributed to a less anxious environment. Even though agile members were experiencing a higher level of anxiety initially because of a lack of knowledge in methodology and project, the analyses revealed that the increased rate in anxiety level was reduced after week 5. From the interview sessions carried out during week 11, it was indicated that the three XP coaching sessions conducted improved the members understanding of *using* story cards, *developing* test cases and *practising* pair programming.

Analysis of the depression variables suggested that XP members have low depression levels in the most dynamic project. One explanation for this finding is that even though the dynamic environment allows only partial requirements to be captured at the early stage, the XP approach allows its members to proceed quickly to the coding and testing stage of the project. The ability to move forward gave the members a sense of control and progress of the project. In addition, the emphasis of several stages in testing, which is done simultaneously with the coding, allows feedback seeking to take place. Feedback seeking occurs when developers purposely monitor their work for performance information and has the advantages of reducing

uncertainty about work performances(Ashford 1989) and also of managing impressions (of oneself or of the client) by presenting team members as competent (Moses et al. 2003). The initial findings indicate that constant testing, pair discussions and client reviews, which result in constant feedback to the team, are considered as treatment for depression. It should be noted that all of the XP teams practised testing vigorously and simultaneously with coding. In comparison, designed-based teams only tested their software vigorously at the end of the project as is to be expected in a traditional approach.

Feeling enthusiastic throughout project development is often a utopian ideal, especially with a volatile project. The ability of the XP approach to cut the project complexity into small stories and solving them in stages according to the client's priority, allowed the XP members to retain their enthusiastic feelings. By presenting complex software as a sequence of small simple stories, the members had a clear measure of successful achievement as well as a retreat possibility to a previous successful step upon failure. Enthusiasm is also maintained through constant unit and integration testing that is conducted throughout the project. Communications amongst team members and with the clients also supported the feeling of enthusiasm.

The ability to discuss complex projects through simple stories and simple design encourages teamwork. By changing pairs constantly, the XP approach created a knowledge-based developer rather than job-based worker. The flexibility associated with the tasks that these developers can perform is likely to be an enhanced value in dynamic environments and less value in stable environments (Lepak et al. 2003). Other factors such as communication skills and personality are vital in modern SE.

7 Limitation

One potential limitation concerns the use of students and the implications for the generalisability of the results. Scholars are rightly cautious about generalizing from student-based studies to working adults. Attempts have been made to carry out this study in an organization but the setup does not permit comparison study of this nature to be carried out.

8 Conclusion

The effect of methodology on developers depends on the particular types of projects that are undertaken. We believe that this study has been able to throw some initial light on the issue that agile methodology are more people-oriented than process-oriented in a more volatile environment. The wellbeing, with exception of contentment, of the developers using this approach seemed to be better only when the requirements are uncertain or volatile. There is a need to carry out more experimental comparison study to validate this finding.

9 Acknowledgement

We thank Carolyn Axtell for the advise on analysing the wellbeing variables and Jarita Duasa for her advise during the analysis of the quantitative data.

10 References

- Agho, A.O., Price, J.L., and Mueller, C.W. "Discriminant validity of measures of job satisfaction, positive affectivity and negative affectivity," *Journal of Occupational and Organizational Psychology* (65), 11December 1991 1992, pp 185-196.
- Ashford, S.J. "Self-assessment in organization: A literature review and integrative model," in: *Research in Organizational Behavior*, C.L. L. and B.M. Staw (eds.), 1989, pp. 133-174.
- Axtell, C., Wall, T., Stride, C., Pepper, K., Clegg, C., Gardner, P., and Bolden, R. "Familiarity breeds content: The impact of exposure to change on employed openness and well-being," *Journal of Occupational and Organizational Psychology* (75:2), June, 2002 2002, pp 217-231.
- Batt, R., and Appelbaum, E. "Work participation in diverse settings: does the form affect the outcome, and if so, who benefits?," *British Journal of Industrial Relations* (33:3) 1995, pp 353-378.
- Beck, K. *Extreme Programming Explained: Embrace Change* Addison-Wesley, NJ, 2000a, p. 190.
- Beck, K. *Extreme Programming Explained: Embrace Change*. Addison-Wesley, 2000b, p. 166.
- Beck, K., and Fowler, M. *Planning Extreme Programming*, (1 ed.) Addison-Wesley, New Jersey, 2001, p. 139.

- Blomqvist, K., Tikkanen, J., and Moller, K. "Partnering in High-Velocity Environment-Is the Interaction Theory Valid Approach," 18th annual IMP Conference, Dijon, France, 2002.
- Chalykoff, J., and Kochan, T. "Computer-aided monitoring: its influence on employee job satisfaction and turnover," *Personnel Psychology* (42:2) 1989, pp 807-834.
- Clegg, C., Axtell, C., Damodaran, L., Farbey, B., Hull, R., Lloyd-Jones, R., Nicholls, J., Sell, R., and Tomlinson, C. "Information technology: a study of performance and the role of human and organisational factors," *Ergonomics* (40) 1997, pp 851-871.
- Francisco, M.J. "Empirical Assessment of Extreme Programming," in: *Dept of Computer Science*, University of Sheffield, Sheffield, 2004, p. 189.
- Frenkel, S., Tam, M., Korczynski, M., and Shire, K. "Beyond bureaucracy? Work organisation in call centres," *International Journal Human Resource Management* (9:6) 1998, pp 957-979.
- Holcombe, M., Cowling, T., and Macias, F. "Towards an Agile Approach to Empirical Software Engineering," Proceeding ESEIW 2003 Workshop on Empirical Studies in Software Engineering, Roman Castles, Italy, 2003a, pp. 37-48.
- Holcombe, M., Gheorge, M., and Macias, F. "Teaching XP for Real: Some Initial Observations and Plans," in: *Extreme Programming Perspectives*, M. Marchesi, G. Succi, D. Wells and L. Willaims (eds.), Pearson Education Inc, Berlin Heidelberg, Germany, 2003b, pp. 251-259.
- Holman, D. "Employee wellbeing in call centres," *Human Resource Management Journal* (12:4), November 2002, pp 35-50.
- Kitchenham, B.A., Pfleeger, S.L., Pickard, S.L., Jones, P.W., Hoaglin, D.C., El-Emam, D.C., and Rosenberg, J. "Preliminary Guidelines for Empirical Research in Software Engineering," *IEEE Transactions on Software Engineering* (28) 2002, pp 721-733.
- Knight, D., and McCabe, D. "What happens when the phone goes wild? Staff, stress and space for escape for in a BPR telephone banking call regime," *Journal of Management Studies* (35:2) 1998, pp 163-194.
- Lepak, D.P., Takeuchi, R., and Snell, S.A. "Employment Flexibility and Firm Performance: Examining the Interaction Effect of Employment Model, Environmental Dynamism and Technological Intensity," *Journal of Management* (29:5), October, 2003 2003, pp 681-703.
- Macias, F., Holcombe, M., and Gheorge, M. "Design-led & Design-less: One Experiments and Two Approaches," *Lecture Notes in Computer Science*. (2675) 2003, pp 394-401.

- Melnik, G., and Maurer, F. "Perceptions of Agile Practices: A Student Survey," the 3rd International Conference on eXtreme Programming and Flexible Process in Software Engineering, Sardinia, Italy, 2002, pp. 241-250.
- Moses, S.E., Valenzi, E.S., and Taggart, W. "Are you hiding from your boss? The Development of a Taxonomy and Instrument to Assess the Feedback Management Behavior of Good and Bad Performer," *Journal of Management* (29:4), August, 2003 2003, pp 467-606.
- Norman, D.A., Ortony, A., and Russell, D.M. "Affect and machine design: Lessons for the development of autonomous machines," *IBM Systems Journal* (42:1) 2003, pp 38-43.
- Sevastos, P., Smith, L., and Cordery, J.L. "Evidence on the reliability and construct validity of Warr's (1990) well-being and mental health measures," *Journal of Occupational and Organizational Psychology* (65), 20 September 1991 1992, pp 33-49.
- Simons, A.J.H. "Object Discovery: A process for developing medium sized applications," Tutorial 14, ECOOP 1998, AITO/ACM, Brussels :, 1998, p. 109.
- Simons, A.J.H. "The Discovery Method at <http://www.dcs.shef.ac.uk/~ajhs/discovery>," 2003.
- Syed-Abdullah, S., Holcombe, M., and Gheorge, M. "Practice makes Perfect," in: *Lecture Notes in Computer Science LNCS 2675*, M. Marchesi and G. Succi (eds.), Springer-Verlag, Berlin Heidelberg, Germany, 2003, pp. 354-356.
- Warr, P.B. "The measurement of well-being and other aspects of mental health.," *Journal of Occupational and Organizational Psychology* (63) 1990, pp 193-201.

Appendix A

WARR'S WELL-BEING MEASURES

Name: _____

Team: _____

Date: _____

This scale consists of a number of words that describe different feelings and emotions. **During the past week**, how much of the time has **your software project** made you feels:

- 1- Very slightly or not at all
- 2- A little
- 3- Moderately
- 4- Quite a bit
- 5- Extremely

(Please tick / the appropriate column for each feeling)

No.	Emotions/Feelings	1	2	3	4	5
1	Tense					
2	Miserable					
3	Depressed					
4	Optimistic					
5	Calm					
6	Relaxed					
7	Worried					
8	Enthusiastic					
9	Anxious					
10	Comfortable					
11	Gloomy					
12	Motivated					

The Grouping in the Warr scale

WARR SCALE

Anxiety	Tense Worried Anxious
Contentment	Calm Relaxed Comfortable
Depression	Miserable Depressed Gloomy
Enthusiasm	Optimistic Enthusiastic Motivated

**Anxiety-Contentment
Depression-Enthusiasm**

Appendix C

XP ACTIVITES TABLE

No	Practices	Activities/Products			
1.	Planning games	Write Stories cards	Write Task cards	Estimate task, cost	Clients write simple test sets
2.	Coding standard	Identify standard	Naming convention		
3.	System metaphor	Common vision	Shared vocabulary	Analogies	Architecture
4.	Simple design	Story cards	X-machine	XX-machine	
5.	Testing	Test cases	Unit testing	Functional testing	Test suite
6.	Refactoring	Early identification with X-machine			
7.	Short/ Frequent release	Continuous review with client	Early feedback from clients	Changes in requirement	
8.	Continuous integration	Incremental integration			
9.	Pair programming	2 person	Swapping pair		
10.	Collective ownership	Repository	Testing partner	Swapping modules	
11.	On-site customer	Write stories	Discuss stories	Immediate feedback	Discuss user interfaces
12.	Forty-hour week (final week)	Less activities during last week	Completed project	Tested project	Integrated software

Appendix D

SOFTWARE HUT – CLIENT’S ASSESSMENT SHEET

This sheet is meant to help in the assessment of the suitability of the delivered system. The mark obtained for each team will contribute 50% towards their final module mark

Group number: _____

Documentation

Presentation	[5 marks]
User Manual	[5 marks]
Installation guide	[5 marks]

Software system

Ease of use	[5 marks]
Error handling	[5 marks]
Understandability (use of appropriate language etc)	[5 marks]
Base Functionality (completeness)	[5 marks]
Innovation (extra features)	[5 marks]
Robustness (correctness – doesn’t crash)	[5 marks]
Happiness with product	[5 marks]

TOTAL _____

Signed:

Appendix E

LECTURER ASESMENT SHEET –SOFTWARE HUT

Marking Scheme (Discovery group)

Group number: _____

Requirement Documentation _____[5 marks]

A set of user stories

A statement of requirements, signed off by the client.

A detailed design for the proposed system, using an appropriate language _____[10 marks]

Completed the test results. _____[10 marks]

Acceptance of test report.

Complete code listing. _____[10 marks]

This must satisfy the coding standards (where appropriate) together with full documentation.

User Documentation. _____[5 marks]

Installation instructions, maintenance guide/manual.

A commentary on the project (maximum 10 sides) containing: _____[10 marks]

A log of the project describing important milestones

A description of the group structure, the roles of individuals and the mechanism for communication used between group members.

A description of the quality control strategy (who did what, when and how the acceptance criteria was defined)

A list any references to the literature used during the project.

An evaluation of the group performance including an allocation of the proportional effort contributed to the project by individuals in the group. this statement must be signed by all members of the group.

Please supply copies of all team meeting minutes, showing responsibilities for actions, the progress and the revisions made to the project plan.

Total process mark: _____ [50 marks]

This mark is combined with the client's mark which is also out of 50 marks

Appendix F

LECTURER ASESMENT SHEET –SOFTWARE HUT

Marking Scheme (XP group)

Group number: _____

- Requirement Documentation _____[5 marks]
A set of user stories. A statement of requirements, signed off by the client.
A detailed specification of test cases for the proposed system, using an appropriate language. _____[5 marks]
- Test management process. _____ [5 marks]
How do you develop and apply the test? Is there any scripts that you need to develop to automate the process? Use of tools.
- Completed the test results. _____[10 marks]
Acceptance of test report.
- Complete code listing. _____[10 marks]
This must satisfy the coding standards (where appropriate) together with full documentation.
- User Documentation. _____[5 marks]
Installation instructions, maintenance guide/manual.
- A commentary on the project (maximum 10 sides) containing: _____[10 marks]
A log of the project describing important milestones
A description of the group structure, the roles of individuals and the mechanism for communication used between group members.
A description of the quality control strategy (who did what, when and how the acceptance criteria was defined)
A list any references to the literature used during the project.
An evaluation of the group performance including an allocation of the proportional effort contributed to the project by individuals in the group. this statement must be signed by all members of the group. evidence of pair programming.
Please supply copies of all team meeting minutes, showing responsibilities for actions, the progress and the revisions made to the project plan.
- Comments.

Total process mark: _____ **[50 marks]**

This mark is combined with the client's mark which is also out of 50 marks