Proceedings of the IASTED International Conference
SOFTWARE ENGINEERING
February 17-19, 2004, Innsbruck, Austria

AN EMPIRICAL INVESTIGATION OF THE IMPORTANT RELATIONSHIP BETWEEN SOFTWARE REVIEW MEETINGS PROCESS AND OUTCOMES

Yuk Kuen Wong, David Wilson
University of Technology, Sydney
Faculty of Information Technology,
PO Box 123, Broadway, NSW 2007, Australia
zoewong@it.uts.edu.au, davidw@it.uts.edu.au

Abstract
The aim of this paper is to examine the important relationships between software review meeting process and outcomes. One hundred and twenty-one industry practitioners participated in this research. The overall results show that software review meeting process has an effect on outcomes. In particular, the results indicate that (1) status effects have an effect on number of defects found, satisfaction and efficiency; (2) discussion quality has an effect on number of defects found, satisfaction and efficiency; (3) teamwork has an effect on number of defects found, satisfaction and efficiency; and (4) communication has an effect on number of defects found and satisfaction, but no effect on efficiency.

Key Words
Software Review Meeting, Process, Outcomes.

1. Introduction
Software review is considered one of the most cost effective methods for defect detection. However, the review meeting process is often treated as a 'black box'. To achieve higher software review performance, it is important to understand the relationship between software review meeting process and outcomes. Recent Wong's exploratory study ([1]) suggested that software review meeting has effect on review outcomes. As a result, an overall objective of this study is to increases our understanding of what are the key process factors affect software review outcomes.

2. Literature
2.1 Software review process
Software review (inspection) was originally introduced by Fagan ([2]). Recent software review empirical studies mainly focus on defect detection process. An effective group meeting should have common goals, share time amongst group members, share group influence evenly, be willing to disagree and have personal say, listen and build on ideas. It is suggested that four critical factors affect a successful meeting. These include communication, teamwork, status effects and discussion quality ([3]). Each factor will be discussed in the following sections.

2.1.1 Communication
Communication is the key element in any type of meetings. Effective communication can achieve better group performance. Different communication mediums have different levels of richness of information transfer ([4]). It is suggested that face-to-face communication has the highest richness of information transfer compared to any other communication medium ([4]). Languages used in a meeting, ease of understanding, self-expression and willingness to communicate are the indicators for evaluating the effectiveness of communications ([4]).

2.1.2 Teamwork
Teamwork often refers the ability and willingness to work with others and involves information sharing and conflict resolution ([3]). Vroom ([5]) claimed that the higher the support from group members, the higher the performance of the group. Cohesiveness is an important factor influencing group members' support. Cohesiveness means members are "involved in a group" and have the "desire to remain in it" ([4]). This creates a sense of belonging in the group and a strong attraction between group members. It has been suggested that high cohesion can result in better group member support and higher performance ([4]).

2.1.3 Status effect
Status effect refers to "attempts by some participants to intimidate others either verbally or non-verbally, use of influence, status or power to force views on others, to inhibit others from participating in the meeting through their behavior, and pressuring others to conform to a particular view" ([3]). One of major reasons the meeting process fails is due to the status effect. The status effect results in intimidation, which can diminish the group members' performance.

2.1.4 Discussion quality
Discussion quality is determined by the meaningfulness, appropriateness, openness, and creativity of the meeting ([3]). The notion of perception is important in discussion

quality. If group members do not perceive a high quality of discussion, then poor outcomes are likely to occur. Davison ([3]) proposed "the generation of novel, creative solutions or ideas is vital because it promotes a reappraisal of the situation. Creative ideas do not only come from individuals; small team of individuals may form to suggest, or at least support, the creation of ideas".

2.2 Software review outcomes

There are two major outcomes from the defect detection process - quantitative and qualitative. At the completion of defect detection, there are two types of quantitative outcomes: the reviewed software artefact, and quantitative outcomes such as defect information recorded in defect forms (e.g. number of defects).

Another common measurement of software review outcomes is efficiency. Efficiency is determined (1) the length of the meeting, (2) the time used for serious discussion, (3) the defects raised in the meeting [3]. Qualitative outcome can be measured by satisfaction the levels of the group members. Satisfaction has been widely measured in small group literature. It is suggested that the more satisfied group members are, the higher the potential for better outcomes ([3]). Since it creates a positive development environment and this potentially can produce better productivity.

3. Research Hypotheses

Further to Wong's preliminary study in 2003 ([1]), it is suggested that status effects and discussion quality are the key factors influencing the number of defects detected. As a result, we formulate:

Hypothesis 1: Status effects (SE) will have an effect on number of defects found (DEF).

Hypotheses 2: Discussion quality (DQ) will have an effect on number of defect found).

In addition, it is proposed that teamwork and communication have significantly effects on efficiency and team members' satisfaction ([1]). Therefore, we have

Hypothesis 3a: Teamwork (TW) will have an effect on efficiency (EFF).

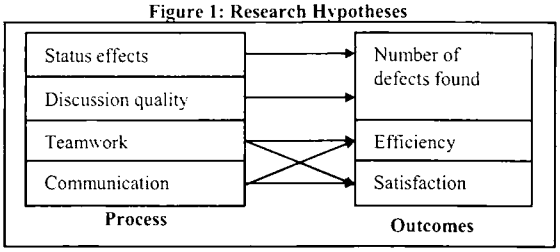
Hypothesis 3b: Teamwork will have an effect on satisfaction (SAT).

Hypothesis 4a: Communications (CM) will have an effect on efficiency.

Hypotheses 4b: Communications will have an effect on satisfaction.

A model shown as Figure 1 presents all the above-formulated hypotheses of the process affect outcomes.

Figure 1: Research Hypotheses



4. Methodology

The objective of this study is to test the two hypotheses of a model of software review, postulating relationships between experience and performance.

To ensure sufficient variance in the data, a random sample of data from companies in six Australian states (i.e. New South Wales (NSW), Victoria (VIC), Queensland (QLD), South Australia (SA), Western Australia (WA) and Tasmania (TAS)) was investigated.

4.1 Questionnaire design

The questionnaire was originally developed by Davison ([3]) and modified by Wong[1] in 2003 ([1]). All questions were five-point-scale and close-ended questions.

4.2 Samples

The main goal of the sampling process chosen was to capture as wide a range of software development companies in the study as possible. Software firms from the computer services category[2] and from the top 500 companies listed in the Australian stock exchange were identified for this research.

The total number of companies selected for the study was 1000. From this a total of 121 individuals voluntarily participated in the survey. Table 1 contains the distribution of the companies sampled in Australia. The following provides a brief review of characteristics of the survey group: A majority of respondents were in age groups ranging between 20 to 60 years old (see Table 2). Approximately 77% of these were male and 23% female (see Table 3). More than half of the participants have university degrees (see Table 4). Approximately 75% are working in computer related occupations (see Table 5).

The number of participants have software industry experience are approximately 93%. About 89% had role experience in requirements review; 91% in design review,

1 The questionnaire can be obtained from Y. K. Wong, zoewong@it.uts.edu.au wongyukkuen@hotmail.com.
2 Categorized size of firms can be based on Australian Bureau of Statistics.

88% in code review and about 72% in testing review. All subjects are industry practitioners and currently working in Australia.

Table 1: Samples distributions

	Frequency	Percent
New South Wales	58	47.9
Victoria	34	28.1
Queensland	10	8.3
Western Australia	13	10.7
South Australia	4	3.3
Tasmania	2	1.7
Total	121	100.0

Table 2: Age

	Frequency	Percent
20 or below	9	7.4
21-30	16	13.2
31-40	54	44.6
41-50	25	20.7
51-60	14	11.6
61 or above	3	2.5
Total	121	100.0

Table 3: Gender

	Frequency	Percent
Male	93	76.9
Female	28	23.1

Table 4: Education

	Frequency	Percent
Doctoral	6	5.0
Master	28	23.1
Bachelor	64	52.9
Undergraduate diploma	10	8.3
Other	13	10.7
Total	121	100.0

Table 5: Position

	Frequency	Percent
Information Technology Manager	28	23.1
System Manager	3	2.5
System Designer	4	3.3
Software Designer	13	10.7
Application and Analyst Programmer	29	24.0
Systems Programmer	3	2.5
Business Analyst	1	.8
Tester	8	6.6
Others	32	26.4
Total	121	100.0

4.3 Measurement Models

In this study, all items were measured on a five-point scale ranging. The measurements of number of defects include:

- True defects (TR) - defects that actually exist and have been successfully detected
- False positive (FA) - defects that do not exist but were wrongly identified
- Net defects (NE) - true defects minus false positive.
- Total issues (TL) - true defects plus false positive.

Note that the measurement of performance is based on the average number of defects found by an individual respondent compared with the average number of defects found by review teams in their company. Individuals were asked to rate themselves in comparing their group members.

The measurements of efficiency include:

- The time given for the software review preparation (EP)
- The time given for the software review examination (EE)
- The meetings results oriented (ER)
- The time spent in the meetings is efficiently used (EU)

The measurements of satisfaction construct include:

- Satisfaction with the overall process of the meeting (SM)
- Satisfaction with the group's performance (SG)
- Satisfaction with the own performance (SO)
- Satisfaction with the role in the software reviews (SR)

The measurements of four meeting process construct include:

- Discussion quality – meaningless vs. meaningful (MM); inappropriate vs. appropriate (AP); closed vs. open (OC) environment setting; familiar vs. creative (FC).
- Communication – language (LA), forwarding ideas (FI), expressing (EX) and understanding (UN).
- Status effects – Power (PO), pressure (PR), intimidate (IN), and inhabitation (IH)
- Teamwork – working together (WT), sharing information (SI), and responsive (RE)

5. Results and Discussions

5.1. Reliability and Validity

In order to validate the content validity, several sources of data were used during the questionnaire development including questionnaire developed by other researchers; The Explicit and Implicit Input-Output "EIIO" model developed from relevant literature but independent from earlier instruments; and feedbacks from pre-tests respondents on the representativeness of questions.

Constructs reliability and validity tests were conducted. Cronbach's ([6]) alphas of discussion quality, communications, status effects and teamwork are between 0.845 and 0.875 respectively (see Table 6). while number of defects found, efficiency and satisfaction are between .8048 and .8849 (see Table 7), which shows good reliability based on Nunnally's Criteria ([7]). A principal components factor analysis with varimax rotation exactly reproduced four factors, explaining 73 percent of the variance indicates both nomological and discriminate validity of the instrument (see Table 8 and 9).

Table 6: Reliability tests – Meeting Process

Constructs	Items	Cronbach alphas
Quality Discussion (PDQ)	Inappropriate vs. appropriate (AP)	.8679
	Meaningless vs. meaningful (MM)	
	Closed vs. open (OC)	
	Familiar vs. creative (FC)	
Communications (PCM)	Expressing (EX)	.8611
	Forwarding ideas (FI)	

Table 6: Reliability tests – Meeting Process

Constructs	Items	Cronbach alphas
	Language (LA)	
	Understanding (UN)	
Status Effects (PSE)	Intimidate (IN)	.8450
	Power (PO)	
	Inhabitation (IH)	
	Pressure (PR)	
Teamwork (PTW)	Working together (WT)	.8751
	Responsive (RE)	
	Sharing information (SI)	

Table 7: Reliability tests – Software Review Outcomes

Number of Defects found (DEF)	True defects (TR)	.8713
	False positive (FA)	
	Net defects (NE)	
	Total issues (TL)	
Efficiency (EFF)	Preparation (EP)	.8048
	Examination (EE)	
	Results oriented ER	
	Efficiently used EU	
Satisfaction (SAT)	Own (SO)	.8849
	Group (SG)	
	Role (SR)	
	Overall process (SM)	

Table 8: Factor analysis – Meeting Process

	PDQ	PCM	PSE	PTW
	F1	F2	F3	F4
AP	0.87			
MM	0.83			
OC	0.82			
FC	0.68			
EX		0.87		
FI		0.82		
LA		0.81		
UN		0.80		
IN			0.82	
PO			0.79	
IH			0.77	
PR			0.75	
WT				0.83
RE				0.82
SI				0.78

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Table 9: Factor analysis – Software Review outcomes

	SAT	DEF	EFF
	F1	F2	F3
SO	0.86		
SG	0.85		
SR	0.82		
SM	0.81		
NE		0.91	
TL		0.88	
TR		0.86	
FA		0.54	
EE			0.82
EP			0.81
ER			0.62
EP			0.52

5.2 Hypotheses Tests

The significance level for all data analysed was 0.05 (two tailed). Pearson's correlation test was used to test the relationships between four process constructs (i.e. communications, discussion quality, status effect and teamwork) and three outcome-constructs (number of defects found, efficiency and satisfaction). Table 10 shows the results of the Pearson's correlation test. The results show that there is a strong negative relationship between status effects and number of defect found ($r = -.243$, $p < 0.01$); there is a strong positive relationship between discussion quality and number of defect found ($r = .467$, $p < 0.01$). These results indicate that hypothesis 1 and 2 are strongly supported. There is also a strong positive relationship between teamwork and outcomes (efficiency: $r = .492$, $p < 0.01$; satisfaction: $r = .528$, $p < 0.01$). These results suggest that hypotheses 3a and 3b are supported.

Interesting results demonstrate that that is no relationship between communication and efficiency ($r = .112$, p= not significant). However, communication is the key driver to group members' satisfaction in software review meetings ($r = .238$, $p < 0.01$). Hence, hypnosis 4a is supported, but rejected hypothesis 4b.

In addition, we found that there is strong relationship between status effect, efficiency and satisfaction (efficiency: $r = -.386$, $p < 0.01$; satisfaction: $r = -.473$, $p < 0.01$); there is discussion quality, efficiency and satisfaction (efficiency: $r = .552$, $p < 0.01$; satisfaction: $r = .365$, $p < 0.01$). On the other hands, teamwork has a positive relationship to number of defect found ($r = .453$, $p < 0.01$); while communications has a relationship with number of defects found ($r = .191$, $p < 0.05$). We also found that status effects have a strongly relationship between communication, teamwork and discussion quality (communication: $r = -.331$, $p < 0.01$; teamwork: $r = -.422$, $p < 0.01$; discussion quality: $r = -.309$, $p < 0.01$); teamwork has a significant relationship between discussion quality ($r = .588$, $p < 0.01$); but there is no relationship between teamwork and communication ($r = .071$, p = not significant).

Although correlation analysis demonstrated the positive relationship between software review process and outcomes, a regression analysis is necessary in order to test the cause-and-effect relationship. Also, because correlation analysis showed status effect, communications, discussion quality and teamwork have a significant relationship with number of defects found, efficiency and satisfaction, we want to know which construct is the most significant determinant of software review outcomes. Hence, we carried out a set of regression analyses: performance on role experience, performance on working experience.

Table 11 shows the regression results. Hypotheses 1, 2, 3a, 3b and 4b are supported. However, result also indicates that communications do not determine efficiency. Figure 2 shows the revised model.

Table 10: Results of correlation analysis on the relationships between process and outcome constructs

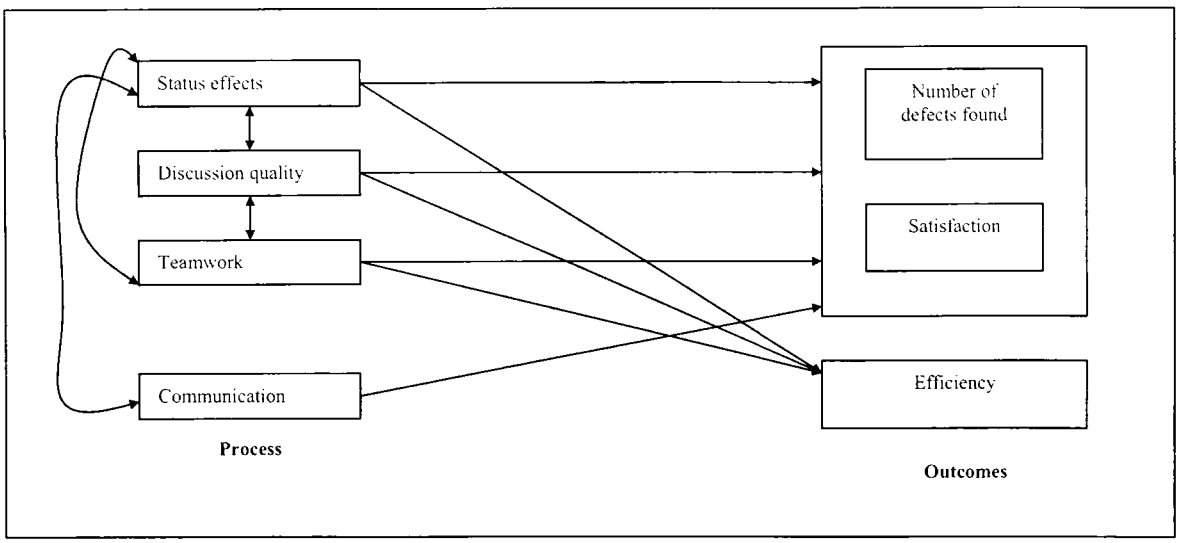
	PCM	PSE	PTW	PDQ	DEF	EFF	SAT
PCM	1						
PSE	-.331**	1					
PTW	.071	-.422**	1				
PDQ	.174	-.309**	.588**	1			
DEF	.191*	-.243**	.453**	.467**	1		
EFF	.112	-.386**	.492**	.552**	.587**	1	
SAT	.238**	-.473**	.528**	.365**	.298**	.471**	1

*** P< 0.01.*
**P < 0.05.*

Table 11: Regression Analysis

	DEF	EFF	SAT
PCM	R=.191* R^2=.036	R=112 R^2=012	R=.238** R^2=.057
PSE	R=.243** R^2=.059	R=.386** R^2=.149	R=.473** R^2=.224
PTW	R=.453** R^2=.206	R=.492** R^2=.242	R=.528** R^2=.280
PDQ	R=.467** R^2=.218	R=.552** R^2=.305	R=.365** R^2=.133

Figure 2: Process affects outcomes



6. Conclusion

The main goal of this study was to validate the relationships between group meeting process and outcomes in the software review process. The main findings of this study are summarized below.

- **H1a accepted: Status effects (SE) will have an effect on number of defects found (DEF).**
- **H2 accepted: Discussion quality (DQ) will have an effect on number of defect found).**
- **H3a accepted: Teamwork (TW) will have an effect on efficiency (EFF).**
- **H3b accepted: Teamwork will have an effect on satisfaction (SAT).**
- **H4a rejected: Communications (CM) will have an effect on efficiency.**
- **H4b accepted: Communications will have an effect on satisfaction.**

The results indicate that discussion quality is a key driver to the number of defect detect. It is suggested that open and familiar meeting setting can detect a higher number of defects in the software review meetings. Teamwork and discussion quality have significant relationships with the outcomes. Interesting results also show that that communication has no effect on efficiency. It can be concluded that all hypotheses are supported except hypothesis 4a. The overall study examined the critical software review meeting process factors affecting the software review outcomes.

References

[1] Y. K. Wong, Successful software reviews meetings, *Supplementary Proceedings of 14th IEEE International Symposium on Software Reliability Engineering*, Denver, Nov. 2003.

[2] M. E. Fagan, Design and code inspections to reduce errors in program development, *IBM System Journal*, *15*(3), 1976, 182-211.

[3] R. Davison, Instrument for measuring meeting success. *Information and Management*, *32*, 1997, 163-176.

[4] L. Trevino, Daft, and Lengel, Understanding managers' media choice, J. Fulk and C. Steinfield eds., *Organizational and Communication Technology* (Newbury Park: California: Sage, 1990).

[5] V. H. Vroom, *Work and Motivation*, (John Wiley & Sons: New York, 1964).

[6] L. J. Cronbach, *Essentials of Psychological Testing*, 3rd Ed., (Irvington: New York, 1970).

[7] J. C. Nunnally, *Psychometric Theory*, (McGraw-Hill: New York, 1978).