The Impact of EMS Support on Inspections; Description of an Experiment

Rini van Solingen
Eindhoven University of Technology / Schlumberger RPS
vansolingen@bladel.rps.slb.com

Michiel van Genuchten
Blenks Groupware / S&P Consulting
mvgenuchten@blenks.com

Rob Kusters
Eindhoven University of Technology
rku@tm.tue.nl

Abstract

Fagan inspections are a structured review of development documents that consists of individual preparation, a meeting and rework by the author of the document. Industrial experience shows that the inspection meeting can be improved by EMS support. The industrial results generate many research questions, relevant for both the EMS and the software engineering community. This paper outlines an experiment with 30 students that aimed to gain insight into the impact of EMS support on the inspection yield. One conclusion is that the results show no improved yield for the EMS supported inspections. The differences between the individuals were much larger than the difference caused by the tools used in the meeting. This can be explained by the fact that the participants were students performing their first inspection. Another conclusion is that the 'paper' inspection created a greater sense of achievement than the EMS inspection, despite the fact the results of both groups were alike.

1. Introduction

Inspections are structured reviews aimed at early defect detection in development documents such as specifications, design and code. Inspections consist of individual preparation followed by a logging meeting. EMS's have been applied successfully to support the logging meeting of inspections in industry [2]. The early industrial results generate many research questions, relevant for both the EMS and the software engineering community. It is in the interest of the software engineering community to improve inspections. It is in the interest of the EMS community to learn from inspections. Due to their quantitative nature, inspections can act as a Guinea pig to study all kind of independent variables that affect the effectiveness of electronic support in meetings. Examples of such questions are:

- What is the impact of the (geographical) location of the participants on the effectiveness of the inspections?
- How is the inspection yield affected by EMS support?
- Are tailored inspection tools more effective than generic EMS tools?
- What is the impact of the distribution of the logging meeting over time?

The industrial environment does not allow doing replicated experiments to investigate the impact of these kinds of independent variables. Therefor a number of experiments have been started at various universities around the world to address research questions such as the ones above. This paper outlines the design of such an experiment and the results of a study with 30 students. The main research question addressed in this study is the second one listed above: How is the inspection yield affected by EMS support?

2. EMS Supported inspections in the field

An inspection aims to detect defects in software development documents such as specifications, design and code. Software engineers around the world have carried out inspections for over twenty years. Key characteristics of inspections are individual preparation, data collection and a fixed syntax to report defects. The basics inspection process consists of:

- a *preparation phase*, during which inspectors are assigned, the document is handed out and the individual preparation is done;
- a *logging phase*, during which the inspectors attend a logging meeting where defects are stated, new defects are

1060-3425/98 $10.00 (C) 1998 IEEE

found, and causal analysis is performed;

- A *rework phase*, during which the document is updated, defects are corrected, and inspection data is stored.

More information on inspections and its application is provided in numerous sources [1, 3, 4, 8].

Like any process, inspections can be improved. The focus of the effort described here is to improve the logging meeting. This meeting has two purposes:

1. consolidating the defects found by the inspectors into one list
2. look for more defects during the logging meeting.

The defects reported by the other inspectors should function as a trigger to detect more defects by the others. The first and second goal of the logging meeting often turns out to be conflicting: inspectors get distracted when other inspectors state a defect. The moderator is usually busy to prevent discussion and to control the logging rate. Inspectors should pick up the pointer to find new defects but they are often also distracted by another minor defect reported. Industry improved the inspection logging meeting by applying an Electronic Meeting System to overcome some of those problems. The original inspection process can be followed in electronic inspections. However, instead of speaking defects out loud, the inspectors submit the defects into a networked computer. Defects appear on the screen of all the other participants.

The initial field study involved the use of an EMS to support the logging meeting of a total of 14 electronic inspections in Philips Medical Systems and Baan Company. The results indicate that the electronic logging meeting contributed much more to the overall result of the inspection than was the case in traditional inspections. The inspections at Baan showed that more defects were found in inspections that were supported by EMS. The results, both at Philips and at Baan showed further that meeting effectiveness (defects per page) and efficiency (defects found per person hour invested) increased considerably. An evaluation of the EMS supported inspections by the participants indicated that the participants were favorable towards the use of the EMS in the inspections. For a more detailed discussion of this field study we refer to [2].

3. Experiment description

The aim of the experiment described in this paper was to investigate the impact of EMS support on inspection yield. The hypothesis was that EMS supported inspections would show a higher yield than the manually supported inspections.

The experiment was conducted with a group of 30 master students as part of a course on information systems design. The students benefited from the experiment in two ways: they experienced the benefits of inspections by doing one themselves and they were introduced to an electronic meeting system in a real life situation. The students had received a 45-minute lecture on inspections the week before the experiment.

The document that was inspected was a four-page chapter out of a draft master thesis. The document described the production flow in a chemical plant. It consisted of flat text and one data flow diagram. We injected two major defects to be sure that some defects were available. We could have saved ourselves the trouble, as will be explained later.

The experiment was conducted in three days over a period of a few weeks. Students were invited for individual preparation of 60 minutes. The students received a complete inspection package:

- an introduction, which described inspections shortly and explained the assignment
- the description of the production flow that was to be inspected (4 pages),
- a high order document from which the document was derived (1 page),
- a standard (1 page),
- a checklist (1 page), and
- a logging form on which defects should be reported.

The preparation took place in one room with six students at a time. The students were not allowed to communicate about the defects they found. The participation took exactly 60 minutes for all groups. All groups took the preparation very seriously and there was hardly any conversation during the preparation. The students were not asked to distinguish severity of defects as is usually done in inspections. The classification of defects into the ‘no defect’, ‘typo’, ‘minor’ and ‘major’ categories was done by the researchers after the experiment.

At the end of the preparation the group of students was split in two groups. Please take notice of the fact that the inspectors were not aware of the type of logging meeting they would be exposed to. The first group went into a manual logging meeting and the second group went into an EMS supported logging meeting in another room. All logging meetings were facilitated by an experienced inspection moderator. A scribe to log the defects further supported the manual logging meeting. The EMS meeting was supported by GroupSystems. The logging meeting took 40 to 60 minutes, depending on the amount of defects found during the meeting. After the meeting evaluation took place by means of an electronic survey and a verbal discussion. All the inspection material and the defect lists were collected after the meeting to prevent one group of students influencing the others.

In total we had invited 12 groups of three students to participate in the experiment. Two groups do not show up in the results. One group did not show up at all. In the second group we encountered a participant with over ten years of experience in software development and extensive experience in inspections. Therefor we excluded this group from the results.

4. Results

4.1 The number of defects found

The number of defects found by the students in the case description was very high: a total of almost 100 defects were found in the four-page document. The number of defects classified as Major was 24. The number of major defects found is realistic. The number of minors found was too big compared to what is usually found in practice [3, 4]. It would therefore have been better to clean up the document before the inspection. The remainder of the results presented will focus on the number of majors found. It must be acknowledged that the large number of minors found may have distracted the inspectors from finding majors.

4.2 Number of defects found in preparation

The number of major defects found during the preparation by the individual inspectors is given in Table 1.

Table 1 Number of defects found during preparation

Group	EMS	Paper
1	10	7
2	11	-
3	6	10
4	-	10
5	8	5
6	7	6
AVERAGE	8,4	7,6

The difference between the number of defects found by both types of logging meetings is small. This is conform expectation because preparation by the members of both groups was identical. The differences between the 10 groups of three students are considerable: the number of defects found in preparation ranges from 5 to 11.

4.3 Number of defects found in the meeting

The number of major defects found in the logging meeting is given in table 2.

Table 2 Number if major defects found in the meeting

Group	EMS	Paper
1	4	8
2	0	-
3	4	2
4	-	1
5	6	4
6	2	2
AVERAGE	3.2	3.4

The number of defects detected ranged from 0 to 8. Again, there is little difference in the number of defects found in the paper and the EMS session. It is obvious that the EMS supported logging meetings do not perform better than the paper logging meeting.

The yield of both inspections is shown in Table 3. Yield is defined as the percentage of the total number of major defects found during the inspections. The total number of major defects was 24. The yield is calculated on the total number of defects found during both preparation and logging meeting. The difference in the yields is small.

Table 3 Inspection yield

% Total Found		
Group	EMS	Paper
1	58%	63%
2	46%	-
3	42%	50%
4	-	46%
5	58%	38%
6	38%	33%
AVERAGE	48%	46%

Looking at the other results, the only considerable difference found is the difference in inspection efficiency. Efficiency in inspections is usually defined as the number of defects found per person hour invested. The number of person hours invested in the EMS inspection is smaller than that invested in the paper inspections due to the fact that the paper inspection required both a scribe (to record all the defects) and a moderator (to manage the inspection). The EMS supported inspection only needed a part-time moderator: besides managing the inspection he also supervised the preparation of the next group of six students. We counted the moderator in the electronic inspection as a half time participant to calculate the person hours invested in the inspection. The resulting efficiency is given in Table 4.

Table 4 Inspection efficiencies

	EMS		Paper	
Group	#faults	fault/ hour	#faults	fault/ hour
1	14	2,5	15	2,0
2	11	1,9	-	-
3	10	1,6	12	1,6
4	-	-	11	1,6
5	14	2,4	9	1,4
6	9	1,7	8	1,2
AVERAGE	11,6	2,0	11,0	1,6

The fact that the EMS system replaces the scribe and makes the moderator available half time to do something else makes the EMS inspection 20 percent more efficient. In practice the moderator would typically use the time that becomes available to act as an inspector himself.

5. The opinion of the participants

Directly after the logging meeting the students filled in a questionnaire. Questions were aimed at background and work experience. Three people had working experience. The fact that one participant had 10 years of software experience was a reason to exclude the results of his group. The other two persons had only one and two years of experience respectively which did not warrant exclusion. Four participants were computer science students. The others were studying either technology management or industrial engineering.

Next we looked at the appreciation of the preparation by means of a number of questions on a 5 point likert scale. Asked if the preparation was effective most participants agreed (22 agree or strongly agree; 7 neutral; 1 disagree). A question that addressed preparation efficiency showed similar results (19 agree or strongly agree; 10 neutral; 1 disagree). No differences between respondents from paper or EMS sessions could be found (as expected).

When we look at the evaluation of the logging meeting a similar view appears (efficiency: 19 agree or strongly agree; 4 neutral; 3 disagree; effectiveness 20 agree or strongly agree; 8 neutral; 2 disagree). However, a difference was noticed between the results from both groups of participants with regard to the perceived effectiveness of the logging meeting. 80% of the paper session participants agreed to the effectiveness of the logging session compared to 50% of the EMS participants. On the other side of the scale none of the paper session participants disagreed with the statement, which compares favourably with the 17% of EMS participants.

This picture is confirmed if we look at the number of errors that the participants think they found during both preparation and logging meeting. No difference in opinion is evident when we look at the results that are claimed based solely on the preparation. When we include the group session a difference appears. Of the paper session 14 out of the 15 participants claim to have found the most important defects whereas only 1 participant disagrees with this statement. This differs distinctly from the results of the EMS session participants where 10 participants agreed to the statement they had found the most important defects, while three were neutral and two disagreed.

From these figures it may be deducted that the paper

session instilled a greater sense of achievement than the EMS session. The yield data does however not confirm this subjective view that more was achieved in the paper sessions.

6. Conclusions

This experiment did not show that inspection logging meetings supported by an EMS are better than paper sessions in terms of inspection yield. The differences between the various groups are much bigger than the differences between the two types of logging meeting. The only way in which the EMS supported meeting did better was in terms of efficiency: it took less people to support the three people inspection group and thus the number of defects found per person hour was higher.

There are a number of possible explanations for the fact that the results from our experiment do not match the results of the empirical work done in industry previously. We want to discuss two possible explanations. The first explanation is the high number of minor defects to be found in the case. Research on inspections indicates that minor defects distract the inspectors from finding the major defects hidden among them [5]. A subsequent experiment should therefore limit the number of minor defects to be found. In inspection terms: the document should be cleaned up before the inspection, as is done in practice.

A second explanation is the fact that the participants were not experienced inspectors, as in industry. The participants were students performing their first inspection. This has probably influenced the result much more than the tools they used. The high differences between the groups and the individuals in the group suggest that this may have been the case. An interesting question for researchers of the impact of group support is how this influences experiments with students who perform a task for the first time. It could be the case that the difference shows up this clearly in our study given the fact that the meeting we investigated is controlled by numbers and as a result, shows difference between groups and individuals very clearly. The difference is also shown this clearly given the fact that inspections are a deductive activity and we knew that we were looking for 24 defects, which allowed to quantify the results of groups and individuals very precisely.

One last conclusion, the paper session created a greater sense of achievement than the EMS session. This looks contradictory to many studies reported upon before. Usually EMS produces better results and the participants clearly recognize this while performing the task [6, 7]. It is an interesting finding in our opinion, that in this case where results are equal, the participants in EMS meetings are apparently less convinced of their achievement.

References

[1] Fagan, M. Advances in software inspections. *IEEE Transactions on Software Engineering, 7*, (July 1986), 741-755.

[2] Genuchten, M. van, Cornelissen, W., van Dijk, C., Supporting inspections with an Electronic Meeting System, *Proceedings of the 30th Annual Hawaiian International Conference on System Sciences*, 1997.

[3] Gilb, T., Graham, D. *Software inspections*. Reading, MA: Addison Wesley, 1993.

[4] Humphrey, W.S. *Managing the software process*. Reading, MA: Addison Wesley, 1989.

[5] Humphrey, W.S. *A discipline for software engineering*. Reading, MA: Addison Wesley, 1994.

[6] Nunamaker, J.F.; Briggs, R.O.; Mittleman, D.D. Lessons from a decade of Group Support Systems Research. *Proceedings of the 29th Annual Hawaiian International Conference on System Sciences*, 1996, pp. 418-427.

[7] Nunamaker, J.F.; Dennis, A.R.; Valacich, J.S.; Vogel, D.R.; George, J.F. Electronic meeting to support group work. *Comunications of the ACM, 7*, (July 1991), 40-61.

[8] Rooijmans, J.; Aerts, H.; Van Genuchten, M. Software quality in consumer electronic products. *IEEE Software*, 1, (January 1996), 55-64.