Making Meetings Work

Michael Ochs
Fraunhofer Institute for Experimental Software Engineering

Rini van Solingen[1]
LogicaCMG Technical Software Engineering

Every one of us has spent many hours, days, maybe even years in meetings. We all have experienced good meetings and bad meetings. Do software engineers spend large portions of their time in meetings? What factors make such meetings successful? This article presents the results of an industrial measurement study conducted to determine why some meetings are successful while other are not.

Software engineering is more than writing and debugging code. Software engineering is basically a *human* activity [1]. Software engineers participate in meetings, discussions, trainings, and other types of social interactions [2]. Software engineers only spend 30 percent of their time working alone. Fifty percent of their working time is spent in groups of two to three people, and the remaining 20 percent in larger groups and travel [3, 4]. This indicates that software engineers spend more than half their time interacting with other people. This is the reason that we investigated the effectiveness of one such interactive activity: meetings.

No literature was found on the total amount of time spent by software engineers in meetings. This article presents the results of an industrial measurement program that investigated the factors that influence the (perceived) quality of meetings in one department of a telecommunications company in Germany [5]. The analysis is based on a data set of 315 registered meetings, collected in a department with about 50 people over a period of one and a half years. For details on the design of this study, see the sidebar "Setting and Design of the Study," page 24. Based on a total of 1,600 labor hours per person per year, the engineers spent more than 6.5 percent of their time in registered meetings.

The study's objective was to determine why some meetings are successful and others not, and to find out how meeting success could be influenced. Although this success is of course difficult to quantify, we defined a binary metric called *meeting result quality* (values: good and bad)[2]. Please note that we asked the participants for their opinion, meaning that we measure *perceived* meeting quality from a participant's viewpoint. The measurement program was set up in a goal-oriented fashion using the goal/question/metrics philosophy [6, 7]. The data are presented in this article along with the following most commonly asked questions with respect to meeting quality.

Figure 1: *Distribution of Result Quality Across Meetings Types*

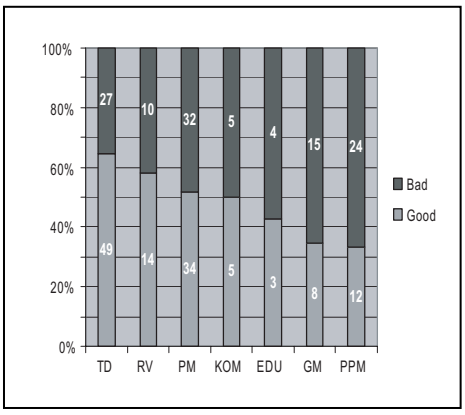
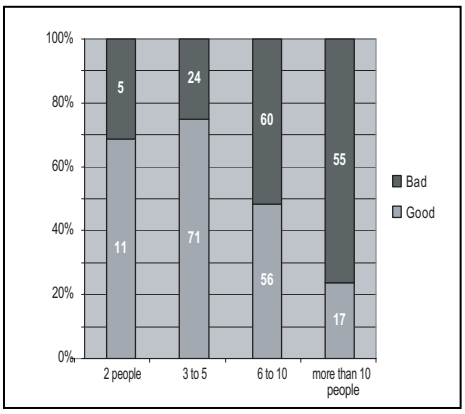


Figure 2: *Distribution of Result Quality Versus Number of Meeting Participants*



Which Meeting Types Are Better?

Several meeting types can be determined, each for different purposes with the goals of the meetings strongly attached to their purposes. The meeting types examined in this study are (with their distribution) as follows:

- **Technical Discussions (TD).** Clarification of technical issues (30 percent).
- **Project Meetings (PM).** Project status and project future direction (28 percent).
- **Project Planning Meetings (PPM).** Project planning (16 percent).
- **Reviews (RV).** Reviews such as code inspections, etc. (10 percent).
- **Group Meetings (GM).** Meetings of organizational units, not project-wise (9 percent).
- **Kick-Off Meetings (KOM).** Initiation of projects (4 percent).
- **Education and Training (EDU).** Training the engineering department, e.g., seminars (3 percent).

The communications culture of the company was very project-driven[3]: 88 percent of the meetings were project related, 12 percent were not project related. The relation between the meeting type and its quality is shown in Figure 1.

A definite correlation between meeting type and its quality could not be found. Figure 1 shows that the probability for a good quality meeting is 35 percent to 65 percent. As such, the meeting type alone does not influence its quality. It seems that the quality of meetings is predominantly determined by other factors than by meeting type.

How Many People Are Present in a Good Meeting?

Meetings take place with a different number of attendees, varying from two persons up to more than 100 persons. In the measurement program, the exact number of participants was measured. For analysis purposes, we divide this number among four categories: meetings of two persons, meetings of three to five persons, meetings of six to 10 persons, and meetings with more than 10 participants. The relation between the number of participants and the percentage of good quality meetings is shown in Figure 2.

In general, the measurements are clear: Good meetings have a limited number of participants. The more people attending, the worse a meeting becomes. Detailed statistical analysis shows that the inflection point lies at eight to 10 partici-

pants, meaning, if a meeting has over 10 participants, it is likely to be perceived as bad. Exceptions were technical discussions, technical review meetings, and group meetings, for which detailed analysis showed that for those, the number of participants does not influence meeting quality.

How Much Speaking Time Is Necessary in a Good Meeting?

As the main purpose of a meeting is interaction, each participant should have sufficient time to explain his or her opinion. Average time per participant is therefore expected to influence how good a meeting is. In the study, the speaking time per participant was estimated from the meeting duration and the number of participants. We grouped these numbers into four categories: below five minutes, five to less than 10 minutes, 10 to less than 20 minutes, and 20 minutes or longer average speaking time per participant. The relation between speaking time and meeting quality is shown in Figure 3.

In order to make the chances for a successful meeting higher, the measurements indicate that at least 15 minutes speaking time is required for each participant. When less time is available, the meeting is likely to fail. Detailed statistical analysis of the measurements showed that speaking time does not influence the quality of review meetings. Group meetings required less time per participant; 10 minutes are sufficient, but this is clarified by the fact that those meetings were mainly used for communicating facts and not so much for discussions.

Of course, the average time per participant is not an exact measure because it is possible that one person speaks, e.g., 50 percent of the meeting time while others do not contribute at all. However, in this study it was not feasible to capture individual participant speaking time in detail. Nevertheless, lessons can be learned from this approximate measure.

How Does the Number of Roles Influence Meetings?

Many roles are present in organizations. Every role has its own responsibilities, goals, and interests. In the measurement program, we measured the influence of the number of roles on the quality of a meeting. Typical roles include developer, project leader, team leader, product manager, quality assurance manager, department head, etc. Figure 4 shows the number of roles present in a meeting and the subsequent meeting quality.

The overall trend shows a decrease in meeting quality with an increase of the number of roles involved. This general trend shows that when more than three roles are present, the worse the meeting. Detailed statistical analysis showed that for technical discussions, reviews, and group meetings, the number of roles had no impact on the meeting's quality.

How Does Hierarchy Affect Meetings?

Besides the number of roles, the amount of hierarchy also was expected to have an influence on meeting quality. Hierarchy was measured by maximum hierarchical distance (XHD), meaning the number of hierarchy layers between the highest and lowest representatives. An XHD of zero means a meeting with peers. An XHD of three means, for example, a meeting in which a software engineer, project leader, department head, and research and development director participate. The relation between hierarchy and meeting quality is shown in Figure 5.

The measurements showed a quite equal distribution when maximum hierarchical distance was zero to two. In those cases, 60 percent of the meetings were good, 40 percent were bad, implying a lower impact of hierarchy on meeting quality. As soon as, however, the XHD was three, the worse the meetings were perceived. Again, this relation was not present for technical discussions, reviews, and group meetings for the same reasons as given before.

How Long Does a Good Meeting Last?

This is a nice and practical question. It would be good to know generally how long meetings should last to make them good. In Figure 6, the relation between meeting duration and quality is shown.

The result is counter-intuitive. You would expect that short meetings are appreciated above long ones. The measurements actually show the opposite: The longer a meeting lasts, the better its quality perception. Even more so, the measurements indicate that meetings are only of good quality when they last at least two hours. What does this mean? Does it mean that only long meetings should be held? Do people like talking? Is it an indicator that this organization has a meeting culture?

Several reasons explain this trend. First, meeting quality is largely influenced by speaking time. Therefore, when many people participate it is better to have a

Figure 3: *Distribution of Result Quality Versus Participant Speaking Time*

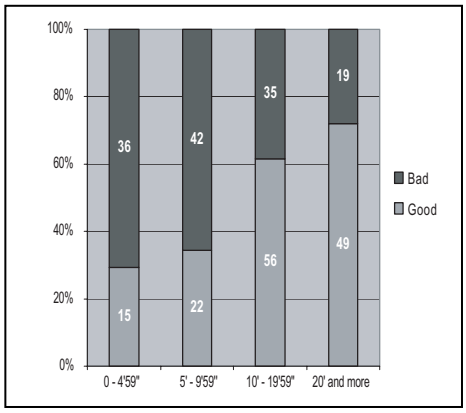


Figure 4: *Distribution of Result Quality Versus Number of Roles*

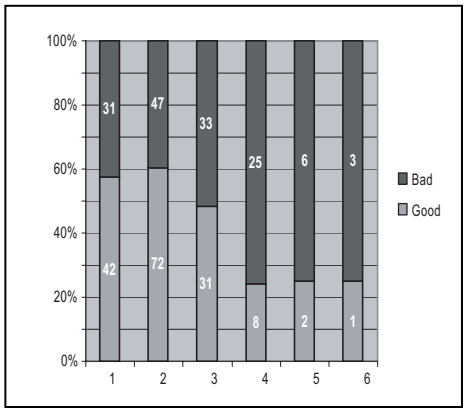


Figure 5: *Distribution of Result Quality Versus Maximum Hierarchical Distance*

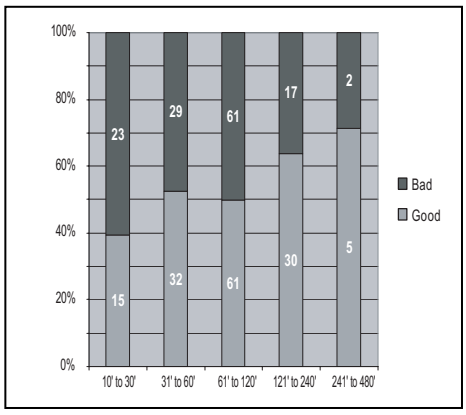
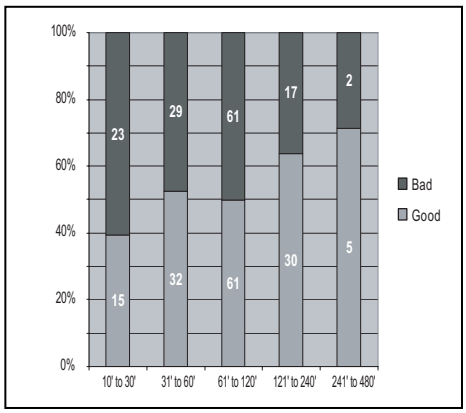


Figure 6: *Distribution of Result Quality Versus Meeting Duration*



Setting and Design of the Study

This study was performed within a goal/question/metric (GQM) [6, 7] measurement-based process improvement program at a telecommunications company's software engineering department, which comprises approximately 50 people engineering control software for phone systems. The time frame was one and a half-years in which 315 communications (i.e., meetings and technical discussions) were registered[4].

Every communication, i.e., meeting, that lasted longer than 10 minutes had to be registered by two data collection sheets. The first one had to be filled in once and was for characterizing the past situation. This was achieved by capturing the type of the communication, the roles involved, the duration, the number of participants, and whether there was moderation or leadership in a communication or not. The second one had to be filled in by every participant measuring the individual perception of the outcome of the situation by the quality of the results of the communication, i.e., the relevance of the communication for the individual. Moreover, the role of every participant was captured.

The goal of the study was to characterize the communication result quality from the perspective of the people working in the particular department. Thus, the variable under study was the perceived quality of results of a communication (Q). The GQM questions were aiming at the relationship between potential factors influencing Q and Q itself, e.g., "How does the number of participants effect the result quality of communications?" "How does the duration of a communication effect the results quality?" More generically spoken, "How does factor X influence Q?" From potentially Q-influencing factors, a set of variables (metrics) was defined, as listed in the metric description table below. Q was measured on a six-point semantic differential scale ranging from satisfactory to unsatisfactory. For further analyses, Q_b was derived from Q by defining the results of a single communication as *bad* if Q was on or above the median of all communications. Overall, the median for Q was *5*, meaning that a total of 50 percent of all meetings was rated 1 through 4, whereas another 50 percent of the meetings was rated *5* or *6* (cf. Q and Q_b in the table below). Thus Q_b can have two values – 0 (bad) and 1 (good). Each participant was asked the following questions to capture his or her perception of the meeting[5]:

I perceive the relevance of the meeting for myself to be …
relevant ❍ ❍ ❍ ❍ ❍ ❍ **non-relevant**

I perceive the quality of the results to be …
good ❍ ❍ ❍ ❍ ❍ ❍ **bad**

I perceive the atmosphere during the meeting to be …
relaxed ❍ ❍ ❍ ❍ ❍ ❍ **tense**

Metric	Description	Scale	Range
TYPE	Type of communication	nominal	TYPE∈{Technical Discussion, Project Meeting, Review, Group Meeting, Kick-Off Meeting, Education, Project Planning Meeting}
LEAD	Type of moderation	nominal	LEAD ∈ {Not Moderated, Moderated}
NOP	Number of participants	absolute	NOP ∈ {2,...,50}
NRI	Number of roles involved	absolute	NRI ∈ {1,...,6}
DUR	Duration of the communication	ratio	DUR ∈ [10,∞)
ATP	Average time per participant	ratio	ATP ∈ (0, ∞)
XHD	Maximum hierarchical distance	absolute	XHD ∈ {1,...,3}
Q	Perceived quality of the results	ordinal	Q ∈ {1,2,3,4,5,6} (1 is worst, 6 is best)
Q_b	Binary quality measure	nominal	Q ∈ {0,1} (0 is lower than Median of Q, 1 is greater than or equal Median of Q, i.e., *bad* or *good* quality)

longer meeting then to compensate on speaking time. Second, shorter meetings are often not planned, meaning that for at least one of the participants it is perceived as an interruption (with the related negative perception) [8]. Finally, one can achieve much better results in a one-day workshop than in eight one-hour meetings [9].

Taking a closer look at the correlation between meeting type and its duration, it is evident there are certain types of meetings where you can say, "The longer the meeting the better it was perceived." Whereas there are also meetings where you can say, "Keep them short to get good results from motivated people." Detailed analysis of meeting types shows the following general visible trends:

- The longer a technical discussion lasts, the better the results are perceived.
- The longer a kick-off meeting lasts, the worse the quality of the results is perceived.
- There is not a clear general trend visible for review meetings.
- For project meetings, no clear trend was identified in the behavior of quality vs. duration; there must be other influencing factors out of the reach of this study.
- The project planning meetings, which are largely perceived to be *bad*, show a switch to a perception of *good* when lasting for at least four hours.
- For group meetings, which are perceived almost the same as the project planning meetings, you can state that a longer group meeting, e.g., of at least two hours, is more likely to be perceived *good* than a shorter one.
- Finally, the analysis shows that seminars and trainings seem to last long and have a comparatively high quality.

Conclusions

Software engineering is more than just writing code. It is a multi-disciplinary job, largely depending on peoples' social skills and social interaction. We emphasize the need for more thorough research on how software engineers interact, and how this interaction can be made more effective.

We showed some factors that influence the participants' perception after attending a software-engineering meeting. Especially the number of participants (maximum eight) and the average speaking time (minimum 15 minutes) are factors that can be controlled in practice in order to steer toward successful meetings. Much research has been performed in the social sciences on human interaction and communication. For example, social science research showed that five to seven persons in a meeting appears an optimum, when more than 12 persons are present a chairman/moderator is necessary, and above 30 participants no dialogues are possible in meetings [10].

Though our findings may not be valid for all kinds of organizations, we are confident that several of the detected success factors are applicable to other organizations as well. These factors can be worked into the following guidelines:

- Keep the number of meeting participants as low as responsibly possible to foster an effective exchange of information.
- With the lowest possible number of participants, plan for sufficient time to give each the possibility to have a valuable contribution to the meeting; "being

effective is much more important than being ultimately efficient" [11].

- Make a clear agenda and improve the meeting process (e.g., using creativity techniques) of poorly performing meetings, i.e., project planning meetings, group meetings.
- Carefully consider the required hierarchical distance of meeting participants; the measurements indicate a drop in perceived meeting quality when maximum hierarchical distance in a meeting is high.
- A lower number of different roles makes a meeting easier to conduct since people speak the *same language*. The chance for stumbling into contradicting interests of different stakeholders is kept low as well. Low variety of roles keeps – on an average – the hierarchical distance low as well.

It is up to the readers to determine whether these guidelines help them make meetings more successful. We hope this article contributes to increasing the effectiveness of software engineering by uncovering some success factors for software engineering meetings that *make meetings work* and by emphasizing that software engineering highly is a people discipline.◆

Acknowledgements

The authors would like to thank Professor Dr. Lionel Briand, Ralf Kalmar, Dr. Martin Verlage, and Kerstin Lünenbürger for their considerable contributions to the research presented in this article. Furthermore, the authors would like to thank Professor Dr. Shari L. Pfleeger, Dr. Robert Glass, and Professor Dr. Dieter Rombach for their valuable comments to earlier versions of this paper.

References

1. Weinberg, G.M. The Psychology of Computer Programming. Van Nostrand Reinhold Computer, 1971.
2. DeMarco, T., and T. Lister. Peopleware: Productive Projects and Teams. New York: Dorset House Publishing, 1987.
3. McCue, G.M. "IBM's Santa Teresa Laboratory – Architectural Design for Program Development." IBM Systems Journal 17.1 (1978): 4-25.
4. Perry, D.E., N.A. Staudemayer, and L.G. Votta. "People, Organizations, and Process Improvement." IEEE Software July 1994: 36-45.
5. Briand, L.C., R. Kempkens, M. Ochs, M. Verlage, and K. Lünenbürger. Modeling the Factors Driving the Quality of Meetings in the Software Development Process. Proc. of the 10th ESCOM Conference, Apr. 27-29, Maastricht, Netherlands. Shaker, 1999: 17-26.
6. Basili, V.R., C. Caldiera, H.D. Rombach, and R.v. Solingen. "Goal/Question/Metric Paradigm." Encyclopedia of Software Engineering 2nd Ed. Vol. 1. J.J. Marciniak, ed., New York: John Wiley & Sons, 2002.
7. van Solingen, R., and E. Berghout. The Goal/Question/Metric Method. London: McGraw-Hill, 1999.
8. van Solingen, R., E. Berghout, and F. van Latum. "Interrupts: Just a Minute Never Is." IEEE Software Sept./Oct. 1998: 97-103.
9. Arthur, L.J. "Quantum Improvements In Software System Quality." Communications of the ACM 40.6 (June 1997): 46-52.
10. Mitchell, T.R., and J.R. Larson. People in Organizations: An Introduction to Organizational Behavior. McGraw-Hill, 1978.
11. DeMarco, T. Slack: Getting Past Burnout, Busywork, and the Myth of Total Efficiency. Broadway Books, 2001.

Notes

1. At the time of this research, van Solingen was head of the Quality and Process Engineering Department at Fraunhofer Institute for Experimental Software Engineering, Kaiserslautern, Germany.
2. Meeting success was measured on a six-point semantic differential scale capturing the participants' perception of the quality of the results of the meeting.
3. The project, which is referred here, was a mixed maintenance and development project for telecommunications software. Overall, there were approximately 50 people involved in the project, which lasted for approximately 18 months.
4. This set of 315 communications enlarges the study in [5], which relied on ~190 meetings.
5. Only results of the question on quality of meeting results are referred to in this study.

About the Authors

Michael Ochs is a business area manager at the Fraunhofer Institute for Experimental Software Engineering. He is responsible for the Banking and Insurance industry. His interests focus on cost/benefit analysis of information technology and software investments, software engineering decision support, and commercial off-the-shelf-based software development. Ochs is an experienced project manager in various domains of industry and publicly funded research. He holds a diploma in mathematics, computer science and economics from the Technical University of Kaiserslautern, Germany.

Fraunhofer Institute for Experimental Software Engineering
Sauerwiesen 6
D-67661
Kaiserslautern, Germany
Phone: +49 6301 707 251
Fax: +49 6301 707 202
E-mail: michael.ochs@iese.fraunhofer.de

Rini van Solingen is a principal consultant at LogicaCMG and a professor in quality management and quality engineering at Drenthe University in The Netherlands. Van Solingen has been a senior quality engineer at Schlumberger RPS and head of the Quality and Process Engineering department at the Fraunhofer Institute for Experimental Software Engineering in Kaiserslautern, Germany. He has published more than 100 titles in journals, conference proceedings, and books, and is author of "The Goal/Question/Metric Method."

LogicaCMG Technical Software Engineering
P.O.Box 8566
NL-3009-AN
Rotterdam, Netherlands
Phone: +31 10 2537 556
Fax: +31 10 2537 033
E-mail: rini.van.solingen@logicacmg.com