Software Effort Estimation: Unstructured Group Discussion as a Method to Reduce Individual Biases

Kjetil Moløkken[i], kjetilmo@simula.no
Magne Jørgensen[ii], magnej@simula.no

Abstract: *The effort of software projects is often estimated, completely or partially, using expert judgement. This estimation process is subject to biases of the expert responsible. Generally, this bias seems to be towards too optimistic estimates regarding the effort needed to complete the project. The degree of bias varies depending on the expert involved, and seems to be connected to both conscious and unconscious decisions. One possible way to reduce this bias towards over-optimism is to combine the judgments of several experts. This paper describes an experiment where experts with different backgrounds combined their estimates through group discussion. Twenty software professionals were asked to provide individual effort estimates of a software development project. Subsequently, they formed five estimation groups, each consisting of four experts. Each of these groups agreed on a project effort estimate through discussion and combination of knowledge. We found that the groups were less optimistic in their estimates than the individual experts. Interestingly, the group discussion-based estimates were closer to the effort used by the actual project than the average individual expert, i.e., the group discussions led to better estimates than a mechanical combination of the individual estimates. The groups' ability to identify more project activities is among the possible explanations for this reduction of bias.*

Keywords: Software development, effort estimation, expert judgment, group processes, expert bias.

1 Introduction

"When you lie about the future, that's called optimism, and it is considered a virtue. Technically speaking you can't "lie" about the future because no one knows what will happen. When you apply this unique brand of optimism (not lying!) at work, that's called forecasting."

- Scott Adams (2002)

Improving the quality of effort estimation is a great challenge of software project management. Across a wide range of software projects, from web-applications to time critical financial or medical systems, poor effort estimation is observed, and the problem seems to increase with project size (Gray, MacDonnell et al. 1999). To account for this, several formal methods and processes to support estimators, like COCOMO II (Boehm, Abts et al. 2000), ANGEL (Shepperd, Shofield et al. 1996) and WEBMO (Reifer 2000) have been described. This paper, however, focuses on *expert* effort estimation. This seems to be the most common estimating method (Jørgensen, In press), and is employed over a wide range of software projects. One of the main problems with expert

[i] Simula Research Laboratory, Norway
[ii] Simula Research Laboratory, Norway

estimation, however, is that it is no better than its participants, and that it is subject to their individual biases (Boehm 1984; Hughes 1996) and political company pressure (Hughes 1996). A possible method to reduce the risk of unwanted influence on the estimates is to use group discussion as an estimation method. This is no new idea, but it seems to have been “forgotten” in the empirical software engineering research. Nonetheless, there is continuously research in other areas on how to combine the opinions of several experts.

We believe that especially new companies with limited history, process improvement policies and resources will benefit from using groups to support their experts in the effort estimation process. These characteristics apply to many web-development companies. What this paper provides, is a controlled experiment on the performance of estimation groups in a web-development company. A typical Web-development project has several characteristics that separate it from traditional software development projects (Reifer 2000). These characteristics include group size, development approach, processes employed, people involved and estimation technologies used. Some of the differences are due to the nature of the technology involved, but much also has to do with company size and maturity. Most web-development companies are small compared with traditional software development, and smaller companies face additional estimation challenges, like access to domain experts and process management (Moses and Clifford 2000).

In our experiment, the twenty individual experts reported project estimates spanning from 220 to 2 286 hours, with an average of 1 088 hours. The five expert groups answers ranged from 1 100 to 2 251 hours. The actual project effort was 2 345 hours, indicating that some, but not all, of the optimistic bias has been eliminated by group discussion. In essence, the results reported in this paper will apply to similar companies and estimation contexts, but the basic idea, to use unstructured group discussion to reduce individual biases, may also be transferable to more traditional development projects and other estimation contexts. The company studied, our experiences from this and other companies, and to what extent the study has external validity, will be discussed in later sections.

The remainder of this paper will be structured in the following sections: (2) Expert and group estimation background, (3) Hypotheses, (4) Research method, (5) Results, (6) Discussion and (7) Conclusion and further research.

2 Expert and group estimation background

Expert judgment is a commonly employed method to estimate software projects. As reported in a review of studies on expert estimation by Jørgensen (In press), several independent surveys (Heemstra and Kusters 1991; Hihn and Habib-Agahi 1991; Paynter 1996; Jørgensen 1997; Hill, Thomas et al. 2000; Kitchenham, Pfleeger et al. 2002) rate it to be the preferred method among professional software developers. Although expert estimation is commonly used, it is probably not because of the preciseness that the method is favoured. In fact, expert estimation seems to be just as imprecise as formal estimation models (Jørgensen, In press). Expert estimates can be especially useful (and often the only option) in companies that lack documented experience from earlier projects (Höst and Wohlin 1998) or have limited estimation resources (Moses and Clifford 2000). This is often the case for young and/or unstable organizations. Both these characteristics apply to most web-development companies, including the one we studied and report from in this paper.

A search for software effort or size estimation in the leading software engineering journals, IEEE Transactions on Software Engineering, Journal of Systems and Software, Journal of Information and Software

Technology and Journal of Empirical Software Engineering, resulted in slightly more than 100 relevant papers. Out of these; only 16 had expert estimation as a topic, and only one of these had group estimation as a topic (Taff, Borchering et al. 1991). Taff and his colleagues present a structured model for expert estimating in groups, called Estimeetings. They do not compare these estimates with individual estimates, only with actual effort. In other words, the leading software engineering papers did not contain a single study on individual versus group-based software effort estimation.

Group expert estimates can, amongst others, be categorized by two different characteristics. The first characteristic concerns the involvement of the estimators. On one side, you can have designated estimation groups (DeMarco 1982), whose only objective is to estimate the project, and do not participate in the development process. The other alternative is that employees who are likely to develop the project are responsible for the estimates. Both approaches have advantages and disadvantages. A separate estimation group may be much less prone to personal or political biases, and more likely to improve their estimation skill over time (DeMarco 1982). On the other hand, to have such a designated team would require a large organization, and good internal communication. Expert estimators, who are likely to implement the solution themselves, will probably get to know the project better than anyone else, and may have a higher motivation for a thorough project analysis (Hughes 1996). For smaller organizations, this approach may be the only possibility due to financial and resource allocation restrictions. A recent study by Jørgensen (2003) found that estimating accuracy improved when the estimator participated in project development, this is also supported in previous research (Lederer and Prasad 1993).

The second characteristic is a structured versus an unstructured approach. An elaborate method to reduce political related problems in groups is to employ the Delphi technique (Helmer 1966), which is often recommended in management papers (Fairley 2002). The Delphi technique does not involve face-to-face discussion, but anonymous expert interaction through several iterations, supervised by a moderator until an agreed-on majority position is attained. A modification of this technique, to include more estimation group interaction, was developed by Barry Boehm and his colleagues, and labelled the Wideband Delphi technique (Boehm 1981). This technique is a hybrid of unstructured groups and the traditional Delphi method. This approach has been suggested as an effort estimation method in books and articles on software metrics (Fenton 1995), software process improvement (Humphrey 1990), project management (Wiegers 2000) and, effort estimation (Hughes 1996).

To our knowledge neither the Estimeetings, the Delphi nor the Wideband Delphi techniques has been subject to empirical research in a software engineering context during the last 25 years. Nor are we aware of any company that employ these methods. On the other hand, we have experienced that many software organizations apply unstructured group discussions in their work leading to software development effort estimates.

3 Hypotheses

Some of the scepticism towards group-based effort estimation may be attributed to misinterpretation of the results from other research areas, such as introductory textbooks in psychology. Many of these are quite extensive in their coverage of the possible dangers of group processes (Brown 1988; Atkinson, Atkinson et al. 1996; Hewstone, Stroebe et al. 1996; Aronson, Wilson et al. 1999; Forsyth 1999). The terms; group polarization and choice shift, are used to describe two similar, often confused, terms in group psychology. These related theories describe, among other aspects, how an initially optimistic decision can be even more optimistic after group discussion. A definition by Zuber, Crott et al (1992) states that choice shift is the difference between the arithmetic average of the individual decisions and the group decision. Group polarization is the difference between individual pre and post group discussion responses. Studies conducted on group decisions have found choice shift and group polarization effects over a range of different areas from burglary to management (Stoner 1961; Wallach, Kogan et al. 1964; Bem, Wallach et al. 1965; Cromwell, Marks et al. 1991; Zuber, Crott et al. 1992).

Much of this knowledge may have been transferred, and simplified, from psychology to professions like project management and general management. These professions often describe the dangers of group interaction, but not *why* or in *which* situation these dangers apply. This may have resulted in an incomplete picture, where valuable aspects like group motivation (Rowe and Wright 2001) and information sharing (Fairley 2002) have been omitted.

We find evidence of this in Software Management textbooks, where both separate estimating groups (DeMarco 1982) and methods like the Wideband Delphi technique (Boehm 1981; Fenton 1995) are described to counter choice shift and group polarization, without giving further explanations of the phenomenon. Boehm (1981) describes possible sources for biased group estimates such as "…group members may be overly influenced by figures of authority or political considerations."

Fortunately, there are research communities that balance this view. According to Kernaghan and Cooke (1990), the engineering management community have gradually accepted that the output of groups will be superior to its average member. The forecasting community are also constantly addressing best practices for combining expert judgements.

Literature review of the Delphi technique in forecasting (Rowe and Wright 2001) suggests that it, on average, outperforms traditional unstructured groups. However, the literature review has also showed that there are tasks where unstructured groups are better suited. In a rich environment, extra information and group motivation may possibly exist in a traditional group, and this can facilitate the process and make it surpass a Delphi group in performance (Rowe and Wright 2001). Maybe typical software estimation processes represent such "rich environments". In an estimation process, there may be several experts who contribute different project experiences and knowledge. Such experiences can more easily be shared in a face-to-face group, than through a moderator. Earlier literature reviews and experiments have concluded that it seems to be less important which combination method, from a set of "meaningful" methods, is used (Fischer 1981). It does, for example, not matter whether simple averaging, unstructured groups, the Delphi technique or other combination methods are used. Other factors, such as cost and political issues, should determine which combination method to employ (Fischer 1981).

Several software engineering textbooks (Boehm 1981; Kitchenham 1996) and papers (Boehm 1984; Hughes 1996; Fairley 2002) point out that forgotten tasks are among the major obstacles of expert estimation, especially when employing a bottom-up estimation approach, i.e., a decomposition of a project into activities and estimation of each activity individually. An estimation group will reduce this problem since several estimators will identify at least as many activities as any one single estimator alone. Especially if you combine estimates from experts with different company roles and experiences, one may be able to include more project activities.

In sum, previous findings lead us to believe that unstructured group collaboration have similar performance compared with structured groups, and that unstructured groups can be used to reduce individual estimation optimism. The latter belief is evaluated through experimentally testing of the following hypotheses:

H1: Group effort estimates are on average less optimistic than the average of the experts' individual effort estimates.

H2: Individual effort estimates are on average less optimistic after group discussion with other experts than before group discussion.

4 Research Method

The research presented in this paper attempts to follow the guidelines suggested by Kitchenham, Pfleeger et al. (2002), which includes specifying as much information as possible about the organization, the participants and the experiments, as well as the complete experimental results. This is done to ensure that the study is easy to replicate, and that the results are appropriately interpreted and transferred.

4.1 The company studied

The company studied is a web-development company, which has the role of an independent contractor. They develop a wide range of complete solutions for their customers. At the time of the study the company had about 70 employees. The employees (except administration and support staff) were allocated to the following four business roles: Engagement Manager/Sales and client responsible (EM), Project Manager (PM), User analyst/Designer (User) and Programmer (Tech). The roles described are similar to those described by (McDonald and Welland 2001). Average participant experience in the IT-business was 6.3 years, in their current role 2.7 years and with effort estimation 1.5 years. Half of the participants had education of at least master level, while the rest had bachelor degrees or comparable.

They have no formal estimation procedure, their projects have short development cycles and the development processes involved are ad-hoc, typical for web-development companies (Reifer 2000; McDonald and Welland 2001). Instead the company bases its estimates on expert judgment. A project estimate is most often made by the person(s) responsible for the project. This is to our knowledge a common situation in web-development companies, and small companies in general (Moses and Clifford 2000).

4.2 The estimation task

Twenty participants were randomly selected from the company. There were five participants from each of the four company roles (EM, PM, User and Tech). Each participant was required to estimate the effort needed to complete a project, based on a requirement specification. This specification, i.e., the document describing the software to be developed, was taken from a project currently under development by the company. The project and the customers were anonymous, and none of the participants had any knowledge about them. The project was development of a "publication solution" for a technical magazine. Its size was medium to large compared to average projects developed by the company. The estimation instructions stated that the experiment participants should behave as realistic as possible, and that the project members were not allocated yet. This was done in order to reduce any political biases (DeMarco 1982; Thomsett 1996). The participants should assume average company productivity for the project. During the sessions, the experimenter was present to answer questions and take notes. The sessions were also video-taped to ensure thorough analyses of the group discussions. The experiment had the following design:

- The participants developed their individual estimates of the effort needed to produce the software during a 45-60 minutes estimation session.
- After the individual session, the participants formed estimation groups, with one EM, PM, Designer and Tech in each group.
- Each group were given about 60 minutes to agree on an estimate for the same project as in the individual estimation session.
- After the group decision, each participant was asked about his personal opinion about how much effort was needed, i.e., how much the group discussions had changed each individuals' opinions about the estimates.

During the experiment each participant answered two questionnaires requiring background information about his experience, education, estimation training and skill level, and his comments to the experiment.

4.3 Validity

The main threat to internal validity is that the experimental setting distorts the realism of the estimation process, and may give invalid results. The following analysis, however, suggests that at least the outcome of the estimation process was realistic. The average project estimate provided by the experiment participants (1 088 hours) did not deviate substantially from the actual effort estimate made by the company (1 240 hours). We must, however, be aware of the large individual deviation, with estimates from 220 to 2 286 hours. There are also validity threats related to the lack of customer contact and limited time available present in our study. For effort estimates applied as input to a bidding process, however, this was a common situation in the company, who developed more than 500 estimates in 2001.

Regarding external validity, the main problem regarding transfer of results to other organizations may be that the participants in the study were from the same company. This was done out of practical reasons, but our experiences with other companies lead us to believe that the company we studied is similar to many other web-development companies, both in size and (lack of formal) estimation process. Our observations of the company also indicate that they are similar to other small companies who employ an ad-hoc estimation approach, as described by Moses and Clifford (2000). It may be a threat to validity that only one project was used. The project chosen may have been especially difficult to estimate. The optimism reduction may not have been of the same magnitude in other projects that were "easier" to estimate.

The impact from the threats to validity means that our findings may mainly be applicable to small companies with lack of formal estimating processes, when estimating more than average "difficult" projects with limited information and strong time restrictions, i.e., estimation situations typical for web-development companies when providing a project bid.

5 Results

The original individual estimates, the group estimates, and the individual opinions after group discussion are displayed in Table 1. There were no participant drop-outs or incomplete responses. All statistical calculations were done with the package MINITAB[i] 13.3 for Windows.

Table 1: Individual pre-group (Before), group and individual post-group (After) estimates.

Business role	Group A		Group B		Group C		Group D		Group E	
	Before	After	Before	After	Before	After	Before	After	Before	After
EM	1 200	1 000	1 550	1 500	1 850	1 500	547	1 000	2 286	2 800
PM	960	1 200	1 820	1 500	300	1 550	914	1 400	984	2 200
User	1 500	1 200	1 140	1 500	1 260	1 500	620	1 000	1 500	2 000
Tech	960	1 000	585	1 400	220	220	660	1 000	900	2 000
Average	**1 155**	**1 100**	**1 273.8**	**1 475**	**907.5**	**1 192.5**	**685.3**	**1 100**	**1 417.5**	**2 250**
Group	**1 100**		**1 500**		**1 550**		**1 339**		**2 251**	
Actual effort	**2 365**									

The group estimates were less optimistic than the average expert opinion in four out of five groups. An analysis was performed with a paired t-test (Wonnacott and Wonnacott 1990), as suggested in similar research on choice shift (Liden, Wayne et al. 1999). Since the hypothesis suggests a direction of effect (groups are less optimistic than average individual experts), the paired t-test was one-sided. The result is displayed in Table 2.

Table 2: Choice shift t-test

	Group average	Individual average	Difference	Pooled StDev	p-value	Size of effect (d)
Estimates	1 548	1 088	460	368	0.024	1.25

We provide the actual p-values, as suggested by Wonnacott and Wonnacott (1990), instead of pre-defining a significance level for rejection. To measure the size of the difference in average values, we included Cohen's size of effect measure (d) (Cohen 1969), where d = (average value group – average value individuals) / pooled standard deviation amongst groups and individuals. The analysis of possible choice shift on the estimates gave a discernible result ($p<=0.05$), and a large effect ($d=>0.8$). Group polarization is, as described earlier, the difference between individual responses made *before* the group session and the responses made *after* the group session (Zuber, Crott et al. 1992). Both median and average individual estimates where less optimistic in the post-group answers compared with the initial responses. The median values increased by 478 hours, from 972 to

[i] http://www.minitab.com

1450. The average individual estimates increased by 336 hours, from 1088 to 1424. A one sided, paired t-test on the before and after values yielded a discernible result (p=0.003), and a medium (d=0.62) effect size (Table 3).

Table 3: Average values, p-values and effect size (d) for a change in individual optimism.

	Average individual estimates after group discussion	Average individual estimates before group discussion	Difference	Pooled StDev	p-value	Size of effect (d)
Estimates	1 424	1 088	336	544	0.003	0,62

6 Discussion

The actual effort of the project was 2 365 work-hours, that is, most participants gave a much too low effort estimate compared with the actual effort. However, the actual effort of a new project based on the same specification may have used less (or more!) effort than the completed project. The actual effort of the completed project cannot, for this reason, be taken as more than an indication of estimation inaccuracy in this experiment. Nevertheless, based on the actual effort and the original estimate, we believe that estimates of less than 1 000 work-hours point at a strong over-optimism. As can be seen in Table 1, 11 out of 20 of the original effort estimates indicated a workload of less than 1 000 hours. There seems to be a tendency that both the group decisions, and the individual post group discussion decisions, where less optimistic than the original estimates. None of the group decisions, and only 1 out of 20 individual post-group estimates, was less than 1 000 hours. The variation among the individual estimates shows that in reality, the opinions of the same project may differ by a magnitude of ten times! The group process is probably more valuable than mechanical combination of estimates because then the experts not only discuss the estimates, but also their assumption behind them (Winkler 1989).

The nature of the task involved in the experiment differs significantly from most previous studies on choice shift and group polarization. Those studies typically ask groups of randomly selected people to decide on choice dilemmas, or professionals in an occupation to determine risk associated with different tasks. Those studies suggest that initially risky or optimistic decisions become more extreme after group discussion. Our study, however, reports the opposite result. One possible explanation is that our groups were able to identify more activities than did our individuals alone. It was explicitly reported, in the questionnaire, by five of the participants in our experiment that forgotten activities were the main reason for their estimates being lower than the group estimates.

The literature on group processes (Brown 1988; Atkinson, Atkinson et al. 1996; Hewstone, Stroebe et al. 1996; Aronson, Wilson et al. 1999; Forsyth 1999) and software management (Boehm 1981; DeMarco 1982; Fenton 1995), frequently warn about the possibility that group pressure (e.g. an unspoken "competition" to appear as the most risky or efficient programmer) and political preferences (e.g. a management demanding optimistic estimates), could lead group decisions to be unreasonably influenced. Observation of post-group opinions that significantly differed from the group responses can indicate that the participants just complied, for example, with an authority. In our experiment, however, the individual post group estimates were much closer to the group decisions than they were to the original individual estimates. This, together with the submitted questionnaires, indicates that the participants may have been mostly influenced by the arguments and extra information about forgotten activities presented in the group.

It is not unlikely that the similar results would have been found if replacing the unstructured group discussions with more structured group processes, e.g., the Delphi wideband technique. The combination of several experts' opinions seems to bee beneficial in most cases. How to combine expert opinions should be decided relative to estimation context factors, such as type of process and project (Fischer 1981). Structured approaches like the Delphi wideband technique, have other qualities than unstructured approaches (Rowe and Wright 2001), and may, for example, be better suited when the level of personal disagreement is high, or political prestige is involved.

In this experiment, none of the participants in each group had the same company role. This may improve the quality of the group (Fairley 2002), as long as they share terminology and understanding of the problem (Helmer 1966). This applies to the participants in our experiment, since they frequently work on the same projects in the same company. Experts with different backgrounds consider the same project from different angles, and are probably able to identify more activities than experts with similar backgrounds. It is possible that groups of experts with similar backgrounds would show less reduction of optimism in their estimates (Jørgensen and Moløkken 2002).

Companies may develop their own Work Breakdown Structure (WBS) and/or estimation checklists to achieve some of the benefits we found from the group discussions, e.g., individual estimators have a lower risk of forgetting activities when using an extensive WBS or high quality checklists. Use of a group collaboration estimation method may therefore be more important when there are especially challenging projects, for example, projects involving new technology and business areas.

7 Conclusions and Further Research

The findings in our study are that group estimates and individual estimates after group discussion were less optimistic and more realistic than the individual estimates before group discussion. The main sources of this increase in realism, seems to be identification of additional activities and an awareness that activities previously identified may be more complicated than the initially appeared. The unstructured group effort estimation approach presented here may be a simple and inexpensive approach for companies to improve their estimating precision.

A topic for further research is to compare the unstructured approach studied in this paper with more structured group processes, such as the Delphi (Helmer 1966) or the Wideband Delphi (Boehm 1981) techniques. Another research topic is to compare unstructured estimating approaches with WBS and checklist-supported estimates.

Acknowledgements

We wish to thank Simula Research Laboratory for funding the experiment, Dag Sjøberg for valuable comments, and all the participants and organizers at the company participating in the study.

References

Adams, S. (2002). Dilbert and the way of the weasel, HarperCollins Publishers Inc.

Aronson, E., T. D. Wilson, et al. (1999). Social Psychology, Addison-Wesley Educational Publishers Inc.

Atkinson, R. L., R. C. Atkinson, et al. (1996). Hilgard's Introduction to Psychology. Orlando, Harcourt Brace College Publishers.

Bem, D. J., M. A. Wallach, et al. (1965). "Group decision making under risk of aversive consequences." Journal of Personality and Social Psychology **1**(5): 453-460.

Boehm, B. W. (1981). Software engineering economics. New Jersey, Prentice-Hall.

Boehm, B. W. (1984). "Software engineering economics." IEEE Transactions on Software Engineering **10**(1): 4-21.

Boehm, B. W., C. Abts, et al. (2000). Software cost estimation with Cocomo II. New Jersey, Prentice-Hall.

Brown, R. (1988). Group Processes. Cambridge, Blackwell Publishers.

Cohen, J. (1969). Statistical power analysis for the behavioral sciences. New York, Academic Press, Inc.

Cromwell, P. F., A. Marks, et al. (1991). "Group effects on decision making by burglars." Psychological reports **69**: 579-588.

DeMarco, T. (1982). Controlling software projects. New York, Yourdon Press.

Fairley, D. (2002). "Making Accurate Estimates." IEEE Software **19**(6): 61-63.

Fenton, N. E. (1995). Software Metrics. London, Thompson Computer Press.

Fischer, G. W. (1981). "When oracles fail--a comparison of four procedures for aggregating subjective probability forecasts." Organizational Behaviour and Human Performance **28**(1): 96-110.

Forsyth, D. R. (1999). Group Dynamics, Wadsworth Publishing Company.

Gray, A., S. MacDonnell, et al. (1999). Factors systematically associated with errors in subjective estimates of software development effort: the stability of expert judgment. Sixth International Software Metrics Symposium, IEEE Comput. Soc, Los Alamitos, CA, USA.

Heemstra, F. J. and R. J. Kusters (1991). "Function point analysis: Evaluation of a software cost estimation model." European Journal of Information Systems **1**(4): 223-237.

Helmer, O. (1966). Social Technology. New York, Basic Books.

Hewstone, M., W. Stroebe, et al. (1996). Introduction to social psychology. Oxford, Blackwell Publishers Ltd.

Hihn, J. and H. Habib-Agahi (1991). Cost estimation of software intensive projects: A survey of current practices. International Conference on Software Engineering, IEEE Comput. Soc. Press, Los Alamitos, CA, USA.

Hill, J., L. C. Thomas, et al. (2000). "Experts' estimates of task durations in software development projects." International Journal of Project Management **18**(1): 13-21.

Hughes, R. T. (1996). "Expert judgement as an estimating method." Information and Software Technology **38**(2): 67-75.

Humphrey, W. S. (1990). Managing the Software Process, Addison-Wesley Publishing Company, Inc.

Höst, M. and C. Wohlin (1998). An experimental study of individual subjective effort estimations and combinations of the estimates. International Conference on Software Engineering, Kyoto, Japan, IEEE Comput. Soc, Los Alamitos, CA, USA.

Jørgensen, M. (1997). An empirical evaluation of the MkII FPA estimation model. Norwegian Informatics Conference, Voss, Norway, Tapir, Oslo.

Jørgensen, M. (2003). An Attempt to Model Software Development Effort Estimation Accuracy and Bias. Submitted to the 7th International Conference on Empirical Assessment of Software Engineering (EASE), Keele, UK.

Jørgensen, M. (In press). "A Review of Studies on Expert Estimation of Software Development Effort." Accepted for publication in Journal of Systems and Software.

Jørgensen, M. and K. Moløkken (2002). Combination of software development effort prediction intervals: Why, when and how? Software Engineering and Knowledge Engineering, Italy.

Kernaghan, J. A. and R. A. Cooke (1990). "Teamwork in planning innovative projects: Improving group performance by rational and interpersonal interventions in group process." IEEE Transactions on Engineering Management **37**(2): 109-116.

Kitchenham, B. (1996). Software Metrics: Measurement for Software Process Improvement. Oxford, NCC Blackwell.

Kitchenham, B., S. L. Pfleeger, et al. (2002). "A case study of maintenance estimation accuracy." To appear in: Journal of Systems and Software.

Kitchenham, B., S. L. Pfleeger, et al. (2002). "Preliminary Guidelines for Empirical Research in Software Engineering." IEEE Transactions on software engineering **28**(8): 721-734.

Lederer, A. L. and J. Prasad (1993). "Information systems software cost estimating: a current assessment." Journal of Information Technology **8**(1): 22-33.

Liden, R. C., S. J. Wayne, et al. (1999). "Management of poor performance: A comparison of manager, group member, and group disciplinary decisions." Journal of Applied Psychology **84**(6): 835-850.

McDonald, A. and R. Welland (2001). Web Engineering in Practice. Proceedings of the Fourth WWW10 Workshop on Web Engineering.

Moses, J. and J. Clifford (2000). Learning how to improve effort estimation in small software development companies. 24th Annual International Computer Software and Applications Conference, Taipei, Taiwan, IEEE Comput. Soc, Los Alamitos, CA, USA.

Paynter, J. (1996). Project estimation using screenflow engineering. International Conference on Software Engineering: Education and Practice, Dunedin, New Zealand, IEEE Comput. Soc. Press, Los Alamitos, CA, USA.

Reifer, D. J. (2000). "Web development: estimating quick-to-market software." IEEE Software **17**(6): 57-64.

Rowe, G. and G. Wright (2001). Expert opinions in forecasting: The role of the Delphi process. Principles of forecasting: A handbook for researchers and practitioners. J. S. Armstrong. Boston, Kluwer Academic Publishers: 125-144.

Shepperd, M., C. Shofield, et al. (1996). Effort estimation using analogy. International Conference on Software Engineering, Berlin, Germany, IEEE Comput. Soc. Press, Los Alamitos, CA, USA.

Stoner, J. A. F. (1961). A comparison of individual and group decisions involving risks.

Taff, L. M., J. W. Borchering, et al. (1991). "Estimeetings: development estimates and a front-end process for a large project." IEEE Transactions on Software Engineering **17**(8): 839-849.

Thomsett, R. (1996). "Double Dummy Spit and other estimating games." American Programmer **9**(6): 16-22.

Wallach, M. A., N. Kogan, et al. (1964). "Diffusion of responsibility and level of risk taking in groups." Journal of abnormal and social psychology **68**(3): 263-274.

Wiegers, K. E. (2000). "Stop Promising Miracles." Software Development Magazine(February).

Winkler, R. L. (1989). "Combining forecasts: A philosophical basis and some current issues." International Journal of Forecasting **5**(4): 605-609.

Wonnacott, T. H. and R. J. Wonnacott (1990). Introductory statistics, John Wiley & Sons, Inc.

Zuber, J. A., H. W. Crott, et al. (1992). "Choice shift and group polarization: An analysis of the status of arguments and social decision schemes." Journal of Personality and Social Psychology **62**(1): 50-61.