The Impact of Irrelevant and Misleading Information on Software Development Effort Estimates: A Randomized Controlled Field Experiment

Magne Jørgensen and Stein Grimstad

Abstract—Studies in laboratory settings report that software development effort estimates can be strongly affected by effort-irrelevant and misleading information. To increase our knowledge about the importance of these effects in field settings, we paid 46 outsourcing companies from various countries to estimate the required effort of the same five software development projects. The companies were allocated randomly to either the original requirement specification or a manipulated version of the original requirement specification. The manipulations were as follows: 1) reduced length of requirement specification with no change of content, 2) information about the low effort spent on the development of the old system to be replaced, 3) information about the client's unrealistic expectations about low cost, and 4) a restriction of a short development period with start up a few months ahead. We found that the effect sizes in the field settings were much smaller than those found for similar manipulations in laboratory settings. Our findings suggest that we should be careful about generalizing to field settings the effect sizes found in laboratory settings. While laboratory settings can be useful to demonstrate the existence of an effect and better understand it, field studies may be needed to study the size and importance of these effects.

Index Terms—Cost estimation, software psychology, requirements/specifications.

1 Introduction

The use of unconscious processes of judgment is essential in most software development effort estimation, either as part of expert judgment-based estimation, so-called "expert estimation," or as part of providing input to estimation models [14]. It is known that these unconscious judgmental processes can be affected, not just by relevant information, but also by information that is irrelevant. Software professionals' effort estimates can, among other things, be strongly influenced by:

- the client's expectation of the effort that will be expended, even when informed that this expectation is not based on any knowledge of the required software development effort [20], [21], [1];
- variation in the wording of the requirements that is irrelevant to the effort, such as describing a maintenance task as "development of new functionality" as opposed to "an extension" [17];
- information that induces wishful thinking, such as information about future opportunities that will be offered, assuming that high productivity is achieved on the project to be estimated [17]; and
- increased amount of information, even when this information has nothing to do with the effort required [10].

In several of the above studies, independent software professionals established that the information was irrelevant for actual effort usage and/or the estimators were explicitly instructed *not* to use that information in their estimation work. In [20], for example, we asked the software professionals whether or not they had used the client's effort expectations as input to their estimates. The software professionals claimed that they did not use this information at all or that it had only a minor impact on their estimates. However, the measured impact of the information was very large. This supports the claim that essential steps in judgmental effort estimation are based on unconscious processes. We discuss the unconscious steps of judgment-based effort estimation in [15]. Similar impacts on judgment from clearly irrelevant information have been found in numerous other professions, e.g., among professionals who are involved in judicial decision making [6] and property pricing decisions [27]. Software professionals clearly do not differ from other people with respect to the extent to which they are influenced by irrelevant information.

The effect that irrelevant and misleading information has on software professionals can have unfortunate consequences. It can, for example, lead to effort estimates that are too low, and hence to loss-inducing bids, project management problems, and low client satisfaction. Therefore, better knowledge about what type of, when, and how irrelevant and misleading information affects effort estimates is essential if software effort estimation is to become more accurate.

- *The authors are with the University of Oslo, and Simula Research Laboratory, PO Box 134, NO-1325 Lysaker, Norway. E-mail: {magnej, steingr}@simula.no.*

Manuscript received 1 Oct. 2008; revised 30 Oct. 2009; accepted 12 Nov. 2009; published online 6 Aug. 2010.
Recommended for acceptance by A. Mockus.
For information on obtaining reprints of this article, please send e-mail to: tse@computer.org, and reference IEEECS Log Number TSE-2008-10-0308.
Digital Object Identifier no. 10.1109/TSE.2010.78.

In this paper, we use the term "estimation-irrelevant information" to denote information that is not causally related to the actual use of effort and consequently should not influence the estimate. This information is frequently relevant for other purposes, e.g., bidding or planning, and may well correlate with the actual use of effort. For example, it is likely that the clients' expectations about cost or the length of a requirement specification correlate with the actual use of effort, although the clients' expectations of low cost or a short requirement specification do not themselves act causally to reduce the actual use of effort. We do not discount the possibility that there are real-world estimation situations in which the information that we term estimation-irrelevant is relevant for the actual use of effort, e.g., situations in which an expectation that the cost will be low, implicitly say something about a client's expectations about the quality of the software. Consequently, the actual relevance of information may be difficult to determine without knowledge of the intentions of whoever is providing the information.

Previous research on the effect of irrelevant information on effort estimates does not include field studies. The fact that this is so provided an important motivation for the field study described in this paper. While it may be reasonable to generalize, to a certain extent, from the results of previous studies to some types of real-life estimation situations, e.g., small maintenance tasks and early project effort estimates that are based on limited information, laboratory-based results are not necessarily relevant for field settings of the type in which carefully selected experts spend several workdays estimating a project. For example, it is possible 1) that the high time pressure on estimation in the laboratory-based studies increased the use of surface indicators that usually correlate with use of effort, e.g., the number of pages in the specification, and 2) that spending more time on the estimation work would lead to the use of a greater number of causal variables. The use of a greater number of causal variables may, in turn, make software professionals better able to resist the influence of irrelevant information. In short, field studies may be required to assess the effect sizes of irrelevant and misleading information in typical effort-estimation field settings.

The remainder of the paper is organized as follows: Section 2 describes the design of the field study. Section 3 describes the results. Section 4 discusses the results and the limitations of the study. Section 5 concludes.

2 Design of the Study

2.1 Research Questions, Motivation, and Hypotheses

Table 1 presents our research questions, the motivations for the research questions, and the hypotheses. The context of all research questions and hypotheses is the estimation of software projects in field settings.

It is also of interest to determine whether the observed effect of irrelevant and misleading information is a result of higher degree of optimism or a simplification/quality reduction of the planned software product or process. We therefore decided to collect information that would enable this type of determination, e.g., a description of the architecture and data on the effort expended on quality assurance and testing tasks, when we observed that the irrelevant and misleading information had a large effect.

As found in previous studies, e.g., [16], a high variation in effort estimates, a skewed distribution, and a few unusual values are to be expected. This means that the mean values are not very reliable as the central value of the distributions. Hence, we based the analysis on differences of median, rather than the mean, for effort estimate of a treatment group. We chose not to remove the highest and/or lowest values since these estimates were based on serious estimation work and were seemingly not of lower quality than the other estimates. We also include analyses based on the log-transformed estimates to enable the use of statistical tests and measures based on normal distribution of values, i.e., ANOVA and Cohen's d, as means to evaluate the robustness of the statistical results.

2.2 Participants

In the second half of 2007, we contacted 240 software companies in Eastern Europe and Asia by e-mail and invited them to estimate five software projects under ordinary payment conditions. The invitation, including the estimation instructions, is provided as the Appendix. Forty-six companies from various countries agreed to complete the estimation work: Russia (15 companies), Ukraine (5 companies), India (7 companies), Pakistan (5 companies), Bulgaria (4 companies), Romania (3 companies), Belarus (2 companies), Moldovia (1 company), Poland (1 company), Serbia (1 company), Slovakia (1 company), and Vietnam (1 company). The companies varied in size from small (fewer than 10 employees) to very large (more than 1,000 employees). All responding companies sent the CVs of the software professionals that would be in charge of the estimation work to document sufficient experience. We accepted only estimators with professional experience from projects similar to those to be estimated, i.e., we allowed only reasonably experienced estimators.

As can be seen from the instructions in the Appendix, we did not inform the companies that their estimates would be used in a research study. This was because we wanted our request to be treated as ordinary estimation work, which it was from the viewpoint of the software companies. We informed the companies that if they did high-quality estimation work for us, we might offer them other opportunities, such as more estimation work. We have already hired for further work several of the companies whose estimation work we assessed to be of high quality. The companies were given all of the information about the estimation work before they accepted it and were paid a fee that reflected their work effort and that was agreed upon by both parties. None of the individual companies can be identified by the reported data and, as far as we can see, the companies could do nothing but benefit from having participated in the study. The companies not only delivered an effort estimate, but also provided descriptions of the proposed architecture, the development platform, essential estimation assumptions, a work break-down of the project, an assessment of the uncertainty of the estimate, and a description of the estimation process they used. The effort spent by the companies on the estimation and estimation related work varied from about 40 to about 100 work hours.

TABLE 1
Research Questions, Motivations and Hypotheses

Research Questions	Motivation	Hypotheses
RQ1: Does the number of pages in a requirement specification affect the effort estimate of the corresponding software projects?	In a student experiment (*unpublished, contact the authors for the data set and the experimental material*) we found that a decrease in the length of a requirement specification from seven pages to one, brought about by such measures as reducing the font size and the wider margins, without changing the content, resulted in a decrease in the mean effort estimate of about 30%. The students spent less than one work-hour estimating the project, the information about the project was limited, and the students were not experienced estimators. Consequently, it is of interest to study the degree to which spending more time on the estimation work, having more information available, and having more estimation experience, would reduce the influence of the estimation-irrelevant surface indicator "length of specification".	**H1:** A reduction in the number of pages in the requirement specification leads to lower effort estimates, even when the written content is exactly the same.
RQ2: When and how much do numerical values (anchors) presented early in the requirement specification that are irrelevant to the actual use of effort affect the estimated effort of software projects?	Numerous laboratory experiments have demonstrated that almost any number presented early in the context of human judgment can work as an estimation anchor; see [6] for an overview. However, there is a need to study the effect size and robustness of these findings in field settings, when more time is spent on the estimation work and when the estimators have a higher level of experience than in the previous studies. The two types of irrelevant numerical values examined in our study were i) the effort used to develop the system to be replaced, and ii) the client's unrealistic expectations of very low cost that were not based on knowledge of the required effort. The effect sizes of this type of estimation anchor in laboratory settings vary a great deal. Unrealistic client expectations typically have the largest effect sizes and are reported to decrease the median estimate by 40% or more; see [18, 21].	**H2a:** Information about the actual, very low effort used to develop the system to be replaced leads to lower effort estimates. **H2b:** Information about the client's unrealistic expectations of very low cost that are not based on knowledge of the required effort leads to lower effort estimates.
RQ3: Does less time available for software development work (shorter development period) and startup several months from now result in lower effort estimates?	Previous studies, e.g., [5, 8], suggest that effort estimation is frequently based on an unfortunate mix of project management concerns (time available) and realistic use of effort (effort estimate). Having less time available should, rationally speaking, seldom lead to less effort spent. The opposite may more frequently be the case. For this reason, estimation models such as COCOMO [2] and SLIM [28] add effort when there is a schedule compression. In a laboratory experiment (see Experiment A in [19]) we found a 40% decrease in median estimates when the software professionals were informed that the software development should be completed within a two-week period, six months from the time of estimation.	**H3:** The information that the client requires the software to be developed in a short period of time, starting six months from now, leads to lower effort estimates.

2.3 Material

The estimation material consisted of five requirement specifications and the estimation instruction in the Appendix. The first requirement specification (S1) was the same for all companies, while the remaining four (S2, S3, S4, and S5) had one original version and one version manipulated with estimation-irrelevant information to test hypotheses H1-H3. S3, S4, and S5 were real-life specifications that had already been implemented by software professionals, while S1 and S2 were specifications developed by students while they were participating in a course on requirements engineering at the San Diego Software Studio. The actual effort, when

TABLE 2
The Five Requirement Specifications

System ID	System domain	Specification length	Specification Manipulation
S1	Dinner reservation.	5 pages	None. All companies received the same specification. This specification was included to enable a control of similarities of groups with different treatments.
S2	Management of appointments at a doctor's office.	12 pages (3 pages)	The text of the two versions of the requirement specification was identical. However, the manipulated requirements specification was reduced to three pages by using smaller fonts, wider margins, and more compressed text. The original version was 12 pages. Test of H1.
S3	Information site for members of an organization.	4 pages	The following text was included early in the manipulated requirement specifications: "*The previous website required about 25 work-hours to build. The estimate of the effort most likely required to build the new website should, however, be independent of this previous use of effort, as the old and the new systems are quite different in size, quality and complexity*." 25 work-hours is a very low value that had nothing to do with the current system. Test of H2a.
S4	An inventory and work order management system for engineering work.	19 pages (5 pages)	The following text was included early in the manipulated requirement specifications: "*The preliminary budget of the new system is $10 000 [corresponding to about 100 work-hours with typical pricing in the country in which it will be built]. The preliminary budget is not built on any knowledge about the actual cost of developing the new system, and will, if needed, be extended to cover the expenses necessary to build a quality system with the desired functionality.*" One hundred work-hours is a very low value for this project and the companies were instructed to not use this as input to their effort estimate. The manipulated requirement specification also had fewer pages (five) compared to the original version (19 pages), but the content was identical. The purpose of combining two manipulations was to try to induce a very large effect, i.e., to show that the effect in field settings can be substantial when more than one optimism inducing elements is used. If the requirement document size manipulation turned out to give no effect (tested by the manipulation in S2), we could use the estimates to study the effect of the budget information, i.e., we could then use the data to test H2b.
S5	Database for information about research studied, integrated with an existing website.	11 pages	The following text was included early in the manipulated requirement specifications: "[the client] *expects that the system development starts February 3, 2008 and can be launched on February 23, 2008. This three-week period should include all development and testing.*" A short development period should, as discussed earlier, lead to the use of the same or more effort, not less. Test of H3.

implementing S3, was approximately 800 work hours. We have no access to the actual effort of S4. The system specified in S5 had been implemented by four different companies. These companies used between approximately 300 and 900 work hours; see [16] for more information about the implementations of this specification. A brief description of the requirement specifications and the manipulations that were introduced is included in Table 2. The main criteria used when selecting the specifications were that the estimation work on each specification should not take more than two days, the specifications should describe web applications of a type such that most of the invited companies would have extensive experience, and the specification should be as complete and precise as other specifications that are typical of its type.

2.4 Data Collection and Analysis Process

The data were collected as follows:

1. First, we searched the Internet for software companies, using free text search on the terms "outsourcing" and "software development." We selected software companies on the basis of the results of this search, inviting only companies whose websites listed relevant experience. The invitation, which we sent by e-mail, included the five requirement specifications and the instructions for performing the estimation work that are included as the Appendix of this paper. Our understanding of what we meant by an effort estimate was described as follows: "*The work we would like you to complete is to estimate the effort (in work hours) your company, with normal quality of the development work, **most likely** would need to complete each of the projects, including testing. An acceptable interpretation of estimated effort (most likely use of work hours) is that it is just as likely that the actual effort overruns as underruns the estimated effort.*" The invited companies were encouraged to ask for more information about the work, if needed. For each of the four requirement specifications S2, S3, S4, and S5, a company was randomly allocated to either the original or the manipulated version. All companies received all specifications. The proportion of companies with positive response to our request was about 20 percent. Note that we allocated the companies to treatment groups before they responded. This, we felt, was necessary to enable the companies to make fully informed decisions about participation. It had, however, the consequence that the number of companies responding to each manipulation could vary.
2. The companies that were interested in the estimation work responded by e-mail and sent us the curriculum vitae of the person(s) supposed to be in charge of the estimation work, the required price, and the date by which they would complete the estimation work. We accepted or rejected the offer using this information as the basis for our decision. The acceptance/rejection of companies to participate was "blind" regarding the manipulations of the specifications received by the companies. This ensured that we could not bias the selection of companies to fit our hypotheses.
3. Then, the companies completed the estimation work for the five projects, in the sequence they wanted, typically within the following two weeks. They sent us questions by e-mail when they needed clarification about the requirements. We tried to respond as similarly as possible to all companies and the responses were made without our knowing whether the company had received the original or the manipulated version of a requirement specification ("blind" responses). The companies sent us their estimation work for approval once they had completed it. Most companies needed one or two revisions before we approved the estimation work. The criteria for approval are described inthe Appendix. About 2,000 e-mails were sent and received during the data collection process.

There are many possible ways of analyzing the differences in the effort estimates of the different manipulation groups. We decided to use more than one measure and statistical analyses. This enabled us to gain some insight into the robustness of the statistical analyses. We include the following analyses: 1) analysis of differences in mean ranks, based on the original effort estimates, using the nonparametric Kruskal-Wallis test, 2) analysis of differences in mean values, based on the log-transformed effort estimates using the parametric test ANOVA. These two analyses were repeated for the S1-adjusted effort estimates. The S1-adjusted estimates of a company were calculated as the original estimate (the estimates of S2, S3, S4, and S5) multiplied by the estimate of S1 divided by the median estimate of all the S1 estimates. The potential benefit of the S1 adjustment is that the variation of the estimates between the companies decreases. One company with consistently high and one company with consistently low estimates would, applying the S1 adjustment, get more similar estimates. The adjustment may, on the other hand, also lead to a decrease in differences in median or mean values. The total effect of the S1 adjustment on the difference in central values and effect sizes is therefore not obvious. The S1-adjusted results should therefore be interpreted carefully.

As measures of effect sizes we use percentage difference in median values and Cohen's d. Cohen's d is only used on the log-transformed values since it is based on the use of mean values. The use of the untransformed estimates would have the consequence that Cohen's d could be dominated by a few large effort estimates.

Our study emphasizes the differences in effect sizes of the manipulation in laboratory and field settings, i.e., to what degree the substantial effect sizes in laboratory settings are found in field settings. One input to the discussion of the trustworthiness of the observed effect sizes is the p-values of the statistical tests. These p-values indicate how likely it is to observe the difference in mean or median values we actually observed (or more extreme differences), given that there are no differences. While this input is important, the p-values should not be treated as the *only* valid input to a discussion of the trustworthiness of the observed effect. This may be especially important when an analysis results in nonsignificant differences, i.e., "high" p-values, and the effects are in the expected direction [12]. The finding of nonsignificant differences cannot, without more analysis, be used to say that the results are inconclusive or that there are no differences. Instead, the results should, in accordance with good meta-analysis principles, be interpreted in the light of previous results and the power of the study. Clearly, when the observed difference is in the same direction as that of previous results in similar contexts, we would find the result more trustworthy than if we found results in the opposite direction of previous results. As an illustration, two nonsignificant results in the same direction may, in a combined analysis, give a significant effect. One challenge when having nonsignificant results is to avoid researcher biases. It may, for example, be the case that we as researchers will tend to

TABLE 3
Effort Estimation Distributions

System	Group	N	Min.	Q1	Md.	Q3	Max.	S1-Md.	Ln-Mean (std)
S1	Control	46	45	119	189	339	1320	189	5.4 (0.7)
S2	Control	22	61	145	330	443	1200	382	5.7 (0.7)
	Manip.	24	116	194	295	434	688	271	5.6 (0.5)
S3	Control	19	160	341	560	658	1836	616	6.3 (0.7)
	Manip.	27	192	300	481	800	2280	597	6.2 (0.6)
S4	Control	23	240	722	954	1382	2461	1018	6.9 (0.6)
	Manip.	23	276	460	724	1158	3371	971	6.7 (0.6)
S5	Control	21	17	171	214	375	1160	448	5.4 (0.7)
	Manip.	24	40	121	142	262	672	96	5.1 (0.8)

believe in "low power" rather than "no effect" to explain nonsignificant findings when motivated to find an effect. In our study, we provide the data (boxplots, effect sizes, p-values, references to previous studies) to support the reader in his/her own interpretation of the trustworthiness of the reported effect sizes in situations with nonsignificant differences. We also, to some extent, add our own interpretation of the trustworthiness of the observed effect sizes in the light of the current and previous studies.

3 Results

3.1 Distributions of Effort Estimates

The effort estimates varied a great deal from company to company, as expected. Table 3 shows the distributions of estimates for the five projects. Q1 is the first quartile, Md (median) is the second quartile, and Q3 is the third quartile. S1-Md is the median of the S1-adjusted estimates. Ln-Mean (std) are the mean and standard deviation of the ln-transformed estimates.

Differences in effort estimates are the results of differences in productivity, use of the development environment, properties of the produced software, and levels of experience and optimism. We have repeatedly found differences in effort estimates in field settings of the size displayed in Table 3 (see, for example [16]).

One company said it did not have the competence to estimate S5 and, for that reason, did not deliver an estimate for that system. As stated earlier, the allocation of treatment before the companies accepted to participate or not to participate made some of the groups larger than the others.

The following sections provide more information about each of the groups' differences in effort estimates. As described earlier, we add an ANOVA-analysis and a Cohen's d based on the log-transformed effort estimation values, and an S1-adjusted analysis, which aimed at a reduction of the variation of the effort estimates.

3.2 Hypothesis 1: Manipulation of the Length of the Specification

Fig. 1 displays a boxplot of the estimates of the manipulated and the nonmanipulated group of companies. Our measures of effect sizes and statistical significance give the following values:

- Analysis based on the original effort estimates: 11 percent decrease in median effort estimates with short specification length. One-sided Kruskal-Wallis gives $p = 0.4$.
- Analysis based on the S1-adjusted effort estimates: 29 percent decrease in median effort estimates with short specification length. One-sided Kruskal-Wallis gives $p = 0.4$.
- Analysis based on the log-transformed effort estimates: Cohen's $d = 0.2$. One-sided ANOVA gives $p = 0.4$.

The data, in spite of the difference in median values, do not give much support to an effect from the manipulation of specification length in field settings. The effect, if it exists, seems to be much lower than the one we found in the previous study with students (see Section 2.1).

3.3 Numerical Anchors

3.3.1 *Hypothesis 2a: Information about the Low Effort of the System to Be Replaced*

Fig. 2 displays a boxplot of the estimates of the manipulated and the nonmanipulated group of companies.

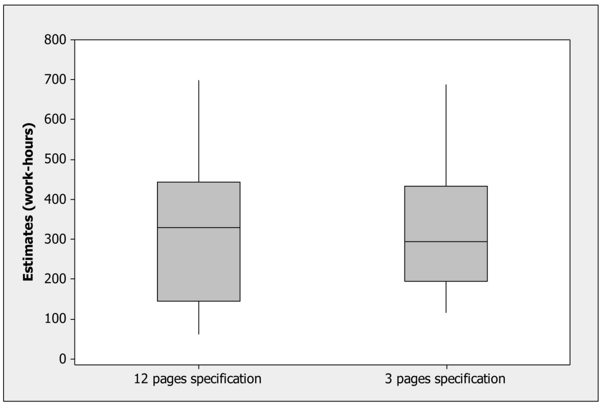

Fig. 1. Boxplot of estimates of S2.

Fig. 2. Boxplot of estimates of S3.

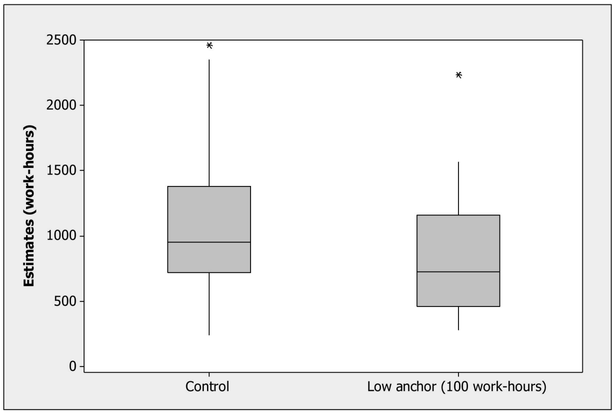

Fig. 3. Boxplot of estimates of S4.

Our measures of effect sizes and statistical significance give the following values:

- Analysis based on the original effort estimates: 14 percent decrease in effort estimates with information about low use of development effort of the previous system. One-sided Kruskal-Wallis, $p = 0.3$.
- Analysis based on the S1-adjusted effort estimates: 3 percent decrease in effort estimates with information about low use of development effort of the previous system. One-sided Kruskal-Wallis, $p = 0.3$.
- Analysis based on the log-transformed effort estimates: Cohen's $d = 0.17$. One-sided ANOVA, $p = 0.5$.

As with S2, the data do not provide much support for an effect in field settings. The effect, if it exists, is lower than those found in similar laboratory-based studies (see Section 2.1).

3.3.2 Hypothesis 2b: Unrealistic Client Budget

As described in Section 3.2, we found that a reduced length of number of pages in a specification had no or almost no effect on the estimates in field settings. We therefore decided to use the manipulation of S4 as a test of the impact that an unrealistic client budget had on the estimates, i.e., we assumed that any observed effect would be due mainly to the client's expectation and not the length of the specification. This expectation on the part of the client was introduced, as described earlier, by providing the information that the client had a preliminary budget representing 100 hours of development work, and that "*The preliminary budget is not built on any knowledge about the actual cost of developing the new system, and will, if needed, be extended to cover the expenses necessary to build a quality system with the desired functionality.*"

The data provide some support for an effect from the above manipulation in field settings (see Fig. 3). Our measures of effect sizes and statistical significance give the following values:

- Analysis based on the original effort estimates: 24 percent decrease in effort estimates with information about low client budget. One-sided Kruskal-Wallis, $p = 0.10$.
- Analysis based on the S1-adjusted effort estimates: 5 percent decrease in effort estimates with information about low client budget. One-sided Kruskal-Wallis, $p = 0.12$.
- Analysis based on the log-transformed effort estimates: Cohen's $d = 0.3$. One-sided ANOVA, $p = 0.13$.

The p-values, together with previous results in the same direction in similar settings, suggest that there is an effect from knowledge about an unrealistically low client budget in field settings. The practical significance is also evident. A 24 percent decrease of estimates, which tend to be over-optimistic without this effect, can be detrimental for the success of a software project. It should, however, be noted that the effects are smaller than the effects from similar manipulations in laboratory settings and that the S1-adjusted difference is much lower.

3.4 Hypothesis 3: Short Development Time with Start-Up Several Months Ahead

The data provide strong support for an effect of the three-week development time with start-up several months ahead on the effort estimate of S5 (see Fig. 4). Our measures of effect sizes and statistical significance give the following values:

- Analysis based on the original effort estimates: 34 percent decrease in effort estimates with short development time with start-up several months ahead. One-sided Kruskal-Wallis, $p = 0.02$.

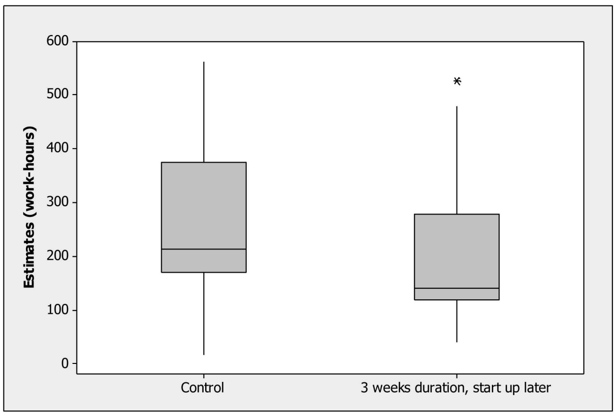

Fig. 4. Boxplot of estimates of S5.

- Analysis based on the S1-adjusted effort estimates: 79 percent decrease in effort estimates with short development time with start-up several months ahead. One-sided Kruskal-Wallis, $p = 0.01$.
- Analysis based on the log-transformed effort estimates: Cohen's $d = 0.4$. One-sided ANOVA, $p = 0.10$.

The data suggest a substantial effect from the manipulation. The actual effort of the four companies that previously developed this system was from about 300 to 900 work hours. Thus, most of the estimates provided by the companies in our study seem to be optimistic.

Above, we argued that it is rational to think that a shorter development period should lead to more, rather than less, use of effort. Yet it may be argued that the information about the schedule is relevant to effort in the sense that a company may want to simplify the solution to make it possible to develop within a short time frame. However, contrary to this, a follow-up analysis of the proposed solutions and the estimation work suggests that it is not likely that the reduced estimates were caused by a simplification of the solutions:

- The specification of S5 was reasonably precise and did not allow much simplification of functionality. (The specification will be sent to interested readers upon request to one of the authors.)
- The average complexity of each of the proposed solutions was assessed by a person with long professional experience as senior architect (the second author of this paper) to be about the same in the two groups. This analysis was "blind," i.e., not based on any knowledge about the group membership.

The companies in the three-week groups may have decided to reduce the quality of the proposed system. An analysis of the proportion of effort that was estimated would be spent on testing and quality assurance suggests that this could, to some extent, have been the case. The median proportion of effort that was estimated to be spent on testing and quality assurance was 13 percent in the three-week group and 21 percent in the control group. The proportion of effort that was estimated would be spent on other types of activities, e.g., project management and design, did not vary much. However, the reduced emphasis on testing and quality assurance is not sufficient to explain more than a small part of the large difference in total effort estimates. Adjusting for the difference in testing effort, the median effort estimates of those companies that received the manipulated version of the specification would only be about 20 work hours higher.

The increased optimism that is found in response to being informed that the development period will be short may be due to the fact that it is easy for companies to think of a three-week development time as indicating a "small" project, and hence, to select reference activities (similar activities that had been completed previously) that correspond to this in use of effort. This bias in the selection of reference activities is in accordance with the "selective accessibility" model described in [25] and discussed in Section 4. Another possible explanation for the increased optimism is that those in the three-week group were told that the development work should start several months ahead, while those in the control group did not receive this information and may have assumed that the development work was supposed to start very soon. It has been found that people may be more optimistic and confident when they are thinking about activities far ahead, rather than about activities that are supposed to be started very soon [7].

The so-called construal level theory may also be useful to understand the estimate decreasing effect of knowing that a project starts several months from now. Construal level theory states that people represent the same information differently depending on whether the information pertains to the near or distant future. People's representations of proximal (near future) information tend to be contextualized and complex (low-level construals), whereas more distal (distant future) information tends to be represented in more abstract and decontextualized terms (high-level construals). The high-level construals are associated with simpler mental models, highlighting the essence of the available information. These construals consist of general, goal relevant features of events. The low-level construals, on the other hand, tend to be richer and more concrete and contain more incidental features of events [31]. At first glance, these ideas seem somewhat intuitive. Because people know less about the far future than the nearer future, it is not surprising that people rely more on abstract and general information when thinking about the distant future. However, the basic tenet of the theory is that people rely on simple mental models when making judgments about the distant future even though they have rich information available. This shift to simpler and more abstract models when making judgment about the distant future may be described as a judgmental heuristic that is learned and is appropriate in most contexts, but also used in (generalized to) situations where it is not appropriate. One deduction from the theory is that people would represent the work needed to create a piece of software with simpler mental models as the start-up time of the project extends into the future. A potential consequence is that people will be more goal and less problem-oriented when the start-up time is distant rather than near. This, in turn, may lead to the lower effort estimates we observed in both laboratory and field settings. We intend to conduct more studies to determine what the explanation is for the strongly increased optimism following this manipulation.

We cannot exclude, in spite of no large observable difference in proposed solution or process, that the companies perceived the schedule and start-up time information as relevant for the use of effort. To some extent, this is what the study results show, i.e., the information did, consciously or unconsciously, impact the effort estimates. What is clear, however, is that the client (us) did not intend this information to be relevant for the use of effort.

4 Discussion of Results

4.1 Comparison of Laboratory and Field Effect Sizes

Table 4 compares the effect sizes in our study with those of the corresponding laboratory studies. We include both the percentage difference in median estimates of the original and Cohen's d based on the log-transformed estimates. We

TABLE 4
Comparison of Effect Sizes of Similar Studies in Laboratory and Field Settings

Treatment	Laboratory settings	Field settings
Shorter specification with same content	30% decrease in median estimates of 181 software engineering students' effort estimates (*unpublished study, contact the authors for the data set and experimental material*) Cohen's d = 0.3	11% decrease in median estimates of S2 (see Section 3.2). Cohen's d = 0.2
Low numerical anchors related to effort of system to be replaced or clients cost expectations.	40% or greater decrease in median estimates from anchors based on, for example, unrealistic client expectations (see for example [18, 21]). Up to about 200 experienced software professionals as participants. Cohen's d of 0.6 and higher.	14% and 24% decrease in median estimates of S3 and S4, respectively (see Section 3.3) Cohen's d = 0.03 and 0.3 for S3 and S4, respectively.
Short development time, combined with start up several months ahead.	40% decrease in median effort (see Experiment A in [19]). 62 software professionals as participants. Cohen's = 0.6	34% decrease in median estimates of S5 (see Section 3.4) Cohen's d = 0.4

have not included the corresponding p-values as we think they are not comparable between the two conditions, e.g., it will be hard to know to what extent the lower p-values found in the laboratory settings are a result of the typically higher number of participants in laboratory studies or are a result of differences in the underlying phenomena.

As can be observed in Table 4, the effect sizes are systematically larger in laboratory settings than in field settings. Although we consider this as the main finding of our study, we also believe that the results should be interpreted with great care. We do not know much about the factors leading to the differences in effect sizes. There may, for example, be small changes in field settings that could lead to similar effects to the laboratory studies. More research on those parts of the differences between typical laboratory and typical field settings are required to gain more confidence in the stability of the observed difference. It should also be noted that the effect sizes in the field setting for S2 and S3 may not be very reliable, i.e., they are based on nonsignificant results. Our data did not provide strong support of the effects of shorter specification and information about the effort of the old system in the field setting.

4.2 Elements of a Model of Judgment-Based Effort Estimation

It may be useful for the improvement of judgment-based effort estimation methods if we could acquire a better understanding of when and why we should expect irrelevant and misleading information in field settings to affect effort estimation. For this purpose, and to better understand the results, we observed in our studies, we explain and apply the elements of a model of human judgment presented in [25].

A basic assumption of the model of human judgment presented in [25] is that human judgment operates by making comparisons. In our context, this means that effort estimation is based on comparisons of the project or task to be estimated (the target) with one or more references, e.g., similar projects or tasks. Adapting the model presented in [25] to software effort estimation contexts, we believe that the essential steps involved in this comparison-based judgment include the following:

1. Selection of a reference for comparison. In software development effort estimation, the references involved in the estimation of a software project can be of different types, depending on the estimation process. When the estimation method is bottom-up, references at the level of activities may dominate the estimation work, while in top-down estimation, the reference may be one or more similar projects. When providing judgment-based input to an estimation model, e.g., when assessing the experience level of the developers or the complexity of the product, a whole range of references may be accessed. Guiding principles for the selection of references are 1) similarity between the reference and target, and 2) the ease with which the reference can be accessed mentally. Projects that are recently experienced or in other ways activated will be more accessible and, as a consequence, more likely input in the judgment-based estimation process.
2. Comparison of the target and reference tasks by searching for information that supports similarity or dissimilarity between the target and reference tasks (similarity/dissimilarity testing). An important, well-documented assumption of the model is that once it has been decided what hypothesis to test, people typically focus on confirming rather than disconfirming evidence [3], [22], [30]. Similarity

testing, rather than dissimilarity testing, may be the dominant mode in most judgmental contexts [25]. Consequently, once a project analogy is found, software developers will tend to search for information that supports the relevance, rather than the irrelevance of this analogy.

3. Synthesis of the collected information about the target and reference tasks in order to estimate the effort that will be required. In [15], we present a model for the synthesis of this information in the context of effort estimation that is based on closeness of analogies and level of uncertainty in the closeness assessment. We also provide empirical evidence in support of the model.

More elements are needed if these model elements are to be used to explain the influence of irrelevant information on the effort estimates. The following well-documented model elements, mainly belonging to Step 1, may be particularly useful for understanding the effects of irrelevant information that we observed in our study:

- *Conversational inference*: A typical, unconscious human response is that information provided in conversational contexts, such as the communication from a client to the provider by means of requirement specifications, is considered to be relevant simply because "*it's there*." Consequently, all types of information that is presented may influence the selection of references, even when it is not relevant for the judgmental task [9], [29], [4]). For example, when the client describes the effort of the previous system in the requirement specification, the software professionals may immediately think that the information is relevant because it has been included in a relevant conversation. Even when they discover that the information is irrelevant for this system, it may not be possible for the professionals to recover an unaffected state of mind. We have documented this problem of returning to an unaffected state in [17].
- *Memory accessibility*: Elements that have nothing to do with the use of software development effort may influence the choice of references. For example, in the study described in [26], a high-price anchor, which was induced by asking the participants whether the average price for a new car is higher or lower than 40,000 German marks (where 40,000 German marks at that time was significantly higher than the actual average price), made it easier to recall expensive cars than less expensive cars. Similarly, estimation-irrelevant information, such as a low preliminary budget may make it easier for the software professionals to recall and use for comparison easy and inexpensive development tasks or projects.
- *Wishful thinking*: The selection of reference projects or tasks may also be influenced by wishful thinking. If the client has indicated that his budget is low, it is not very comforting to suggest an effort estimate that deviates very much from that budget, even when the budget is unrealistic and the software professionals have been informed that it will be extended if necessary. A high estimate may, for example, make the client believe that the reason for this high estimate is low competence. In order to preserve a good image of one's own abilities, a software professional may select reference projects and tasks in accordance with an unconscious intention to avoid negative responses to the estimate on the part of the client. For example, the selection of small and simple reference projects in the S5 project may be a result of the wish to complete the project within the three-week period using only one developer.
- *Categorical references and surface features:* In estimation settings in which little information is available, the estimators have little experience, and/or there is high time pressure (as is the case in many laboratory studies), categorical references and surface features may increase in relevance for the estimation work. This means that information that typically correlates with, but is not causally related to, use of effort is more likely into influence the effort estimate in laboratory settings. For example, if the time that is allocated for the estimation work does not allow the project to be broken down into activities and the estimators do not recall the completion effort of similar projects, it is understandable that the estimators may be influenced by their knowledge that, for example, medium-sized projects typically have requirement specifications that are the same length as the project that they are estimating. It is understandable in light of both the availability of the feature in both the reference projects and the project that they are estimating and the ease with which the feature similarity can be evaluated.

The above model elements explain why irrelevant information may affect the effort estimates. The model also explains why the effect of irrelevant information may be "diluted" in field contexts, in which there is less time pressure, greater expertise, and a greater amount of relevant information about the task compared with laboratory contexts. Less time pressure, greater expertise, and a greater amount of relevant information about the task means that a single piece of irrelevant, misleading information most likely receives less attention and is, for example, less likely to lead to greater accessibility of conversationally induced references. The model also explains why more time, greater expertise, and a greater amount of relevant information do not guarantee a small effect from irrelevant or misleading information. For example, if information about an expectation of low cost on the part of the client has led to the selection of small reference projects, the use of more time on the estimation work may lead to higher confidence instead of more realistic estimates due to search for confirming rather than disconfirming evidence.

4.3 Limitations of the Study

The high number of estimation experiments and experiments in other domains make it safe to claim that it is easy to create situations in which information intended to be irrelevant for the task at hand affects human judgment strongly. That being so, the issue that we address is not whether or not estimation-irrelevant information is able to influence effort estimates, but the degree to which this is likely to occur in field settings where a software company allocates experienced software professionals to estimation work that lasts for several days. The results of our study indicate that the effects may still be present, although typically not as large as in estimation situations in which time pressure is greater, there is less information about the project to be developed, and it is less important to deliver an accurate estimate, i.e., typical laboratory study settings. However, our study has a number of limitations that mean that the results should be interpreted with care.

4.3.1 Pricing or Estimating

We repeatedly instructed the organizations to estimate the most likely use of effort to the client, and not the price. In order to emphasize further that we did not wish to receive a price, we told them that they would not be chosen to develop the projects that they estimated. This was done to avoid a pricing process in which a preliminary budget may be defended as relevant input. However, we observed that despite our instructions, six of the companies were poor at separating these two concerns. Our observation was mainly based on the fact that the effort estimate was used to price the project, rather than estimate the most likely effort. This is consistent with the previous observations of a conflation of concerns about pricing, planning, and estimation among software companies reported in [20]. It is possible to defend the use of some of the effort-irrelevant information if a company conflates concerns about estimation and pricing in its estimation work. For example, a low preliminary budget may be interpreted by the estimator as an indication that low cost is a high priority for client. While we, as clients, intended no such thing, and while we instructed the companies to ignore such information, we cannot fully exclude the possibility that more of the companies conflated pricing and estimation. We conducted the analyses without the six companies that seemed to have provided a price, rather than estimated the effort, and found that for all systems except S5 (the schedule compression manipulation) this resulted in the manipulations having slightly less effect. This weakly indicates that a pricing process may be even more vulnerable to the influence of irrelevant information than processes that emphasize the estimation of most likely effort. A follow-up study on the difference in effect between estimation and pricing settings is recommended.

4.3.2 Unusual Estimation Situation

The typical estimation situation of software development companies is to estimate the effort of projects for bidding, budgeting, or planning purposes. In our study, they provided only the estimate of most likely effort, but they had no opportunity to develop the project. On the one hand, this may have led to less "wishful thinking" (and consequently, the irrelevant information may have had less effect!), but on the other hand, may have resulted in the use of estimation processes different from those that are usually followed. Fortunately, an analysis of the estimation processes used by the companies suggests that the processes were no different from those reported to be in typical use in the software industry (see, e.g., [11], [13]). As an illustration, 41 companies applied a bottom-up, expert judgment-based effort estimation process and five used formal models for effort estimation. There were some variations in the use of judgment-based estimation processes. For example, two companies combined expert judgment with simple rules-of-thumb and five applied the PERT method of minimum-maximum effort estimates. All companies produced proper structures for activities, which was a minimum requirement on our part for accepting the estimation work. As stated earlier, we informed the companies that they would be evaluated and would be more likely to get more paid work from us on the basis of the quality of the estimation work, not on the effort estimate itself. Note that our emphasis on the quality of the estimation work in the instruction and not, for example, the price is likely to result in irrelevant information having less impact than in bidding processes due to greater use of deep-structure indicators than surface indicators as input to the estimation process. A follow-up study on the difference in effect between situation with low and high priority on quality is recommended.

4.3.3 Biased Selection of Companies

Outsourcing companies in low-cost countries, such as India and Russia, were selected for this study because they were less expensive than, for example, Norwegian or North American companies. Within the segment of low-cost outsourcing companies, it is possible that those that accepted the work and payment conditions (about 20 percent of the invited companies) were not representative of companies in those countries or of companies in other regions. They may, for example, spend less time on quality assurance and have less expertise and, as a consequence, be affected by the irrelevant information more easily. Unfortunately, there are no data available to show whether or not there is a bias in one direction or the other and we cannot exclude this possibility. However, due to the fact that we accepted only estimators who had relevant experience and emphasized that we would evaluate the quality of the estimation process, it is also possible that the estimation processes and skills were better than average. Without better knowledge about the world-wide population of software developers and companies, it is difficult to know to what degree a nonrandom sample of companies represents the underlying population. So far, we have seen no survey of software effort estimation that selected software companies by random sampling, other than in national contexts.

4.3.4 Low Commitment

The software professionals were paid for their work and they knew that high quality estimation work might result in their receiving more work from us as clients. However, it is nevertheless possible that some of the companies were not committed strongly to delivering proper estimation work because they knew that they would not have the opportunity to develop the projects that they estimated. However, our observations led us to interpret only one company as not aiming at serious estimation work and we excluded that company from the analysis. The quality of the estimation work was increased by our own quality assurance and the instructions to describe estimation assumptions, technical architecture, development environment, a work breakdown structure, and the estimation process that they used.

4.3.5 What About Estimation Pessimism?

The previous laboratory experiments, such as [20], [1], examine impacts leading to both higher and lower effort estimates. Our field experiments, on the other hand, only examine impacts potentially leading to lower effort estimates. The results of our experiments are consequently limited to only the optimism, and not the pessimism, inducing irrelevant estimation information.

5 Conclusion

Previous studies on the effect of estimation-irrelevant information on the effort estimates have systematically shown large effect sizes. Our study is, as far as we know, the first to investigate the effect sizes in a controlled field setting. We find that the field setting typically led to irrelevant information having smaller impact on the effort estimates than in the more artificial experimental settings. The effect sizes in field settings were not only smaller, but the differences were in the situations with low effect sizes based on statistically nonsignificant results. We do, however, also find that the effects in field setting can be substantial and of practical importance. In particular, we found that unrealistic budget information and receiving the information that the system had to be developed in a three-week period starting a specified date several months ahead in time produced effort estimates that were much lower than those produced by a control group that did not receive this information. We propose elements of a model that may explain the difference in effect sizes between laboratory and field settings.

A possible consequence of our findings is that researchers should be more aware of the different roles of laboratory and field experiments. While a meaningful role of laboratory experiments is to demonstrate the existence of an effect and understand its nature, we should be careful to base statements about the effect size, i.e., the importance of an effect on laboratory studies alone. For the purpose of establishing knowledge about the importance of an effect, we need field studies.

Appendix

Request for Estimates of Software Development Effort

Simula Research Laboratory is responsible for the collection of independent software development effort estimates on five software projects. The independent effort estimates will be applied to evaluate the realism of other software providers' effort estimates of the projects. The five project requirement specifications are enclosed with this mail. The specifications are the clients' original ones, with the exception that we have removed information that identifies the software clients' names. Any requests for more information and clarifications should be directed per email to Dr. Magne Jørgensen at Simula Research Laboratory (magnej@simula.no).

The work we would like you to complete is to estimate the effort (in work hours) your company, with normal quality of the development work, **most likely** would need to complete each of the projects, including testing. An acceptable interpretation of estimated effort (most likely use of work hours) is that it is just as likely that the actual effort overruns as underruns the estimated effort.

Your company's role is consequently **not** to provide a bid for the project, but to help our clients to assess the realism of the effort estimates from other software providers. To ensure that your effort estimates are not impacted by estimation behavior impacted by the wish to win a bidding round (which has repeatedly been found to lead to overoptimistic effort estimates) your company will **not** be considered for the development of the software specified in the specifications.

Your estimation work should be based on the assumption that a tailor-made system has to be developed, i.e., **you should not base the estimates on solutions where you adapt existing software packages or where you base the solution on large premade components.** All important technology choices you make to estimate the effort should, of course, be documented. An important reason for this avoidance of premade packages is the clients' need for control and ownership of source code. Notice also that it is essential that the projects are estimated independently from each others. Estimation work of high quality may also give you the opportunity to conduct similar and other types of software work for Simula Research Laboratory.

For each project we require that:

- The estimation work is completed by software professionals with documented, relevant competence.
- You document the choice of technology and architecture.
- You document essential assumptions you have based the estimates on. This is, in particular, important when the specification is incomplete or difficult to interpret.
- The effort estimates are documented through a work break-down structure, a set of user stories, or a set of use cases. Each element, e.g., each activity, should be estimated.
- The uncertainty of the total effort estimate is documented as specified by us in <*xxx*>.
- You describe the estimation process that you have used.

If this estimation work is of interest to you, please respond to this mail with your price for the estimation work, the CV of the person(s) that will conduct the estimation work, and information about when you will be able to complete the assignment. The CVs should emphasize documentation on estimation experience of projects similar to those included in this mail and be sent to us as soon as possible. The remaining steps are then as follows:

1. We will evaluate whether we choose to use your estimation competence or not. This will be based on your price for the estimation work, the CV, and information about when you will be able to complete the assignment.
2. We will notify you about our decision within one week after reception of the CV.

3. The person(s) described in the CV should then provide the estimates of the most likely use of effort of each of the five projects, as specified in this mail. This should be completed as soon as possible, latest at the date specified by you in the initial mail.
4. When the estimation work is accepted, we pay you for the work. (Or, if you prefer, we pay you 50 percent of the price after the first accepted estimate and the rest when all five are accepted.)

Please consider the requirement specification and the estimates you produce as confidential information, i.e., it is essential that this information is not distributed to people outside your company.

Acknowledgments

The authors thank all of the people who contributed to this study, in particular, Wenche Angel, Hans Christian Benestad, Eric Busboom, Mike Cohn, Tanja Gruschke, Hege Johnsrud, and Dag Sjøberg.

References

[1] J. Aranda and S.M. Easterbroook, "Anchoring and Adjustment in Software Estimation," *European Software Eng. Conf./ACM SIGSOFT Symp. Foundations of Software Eng.,* 2005.
[2] B. Boehm, C. Abts, A.W. Brown, S. Chulani, B.K. Clark, E. Horowitz, R. Madachy, D. Reifer, and B. Steece, *Software Cost Estimation with Cocomo II.* Prentice-Hall, 2000.
[3] B. Brehmer, "In One Word: Not from Experience," *Acta Psychologica,* vol. 45, pp 223-241, 1980.
[4] E. Cowley, "Processing Exaggerated Advertising Claims," *J. Business Research,* vol. 59, pp. 728-734, 2006.
[5] J.S. Edwards and T.T. Moores, "A Conflict between the Use of Estimating and Planning Tools in the Management of Information Systems," *European J. Information Systems,* vol. 3, no. 2, pp. 139-147, 1994.
[6] B. Englich, T. Mussweiler, and F. Strack, "Playing Dice with Criminal Sentences: The Influence of Irrelevant Anchors on Experts' Judicial Decision Making," *Personality and Social Psychology Bull.,* vol. 32, no. 2, pp. 188-200, 2006.
[7] T. Gilovich, M. Kerr, and M.V. Medvec, "Effect of Temporal Perspective on Subjective Confidence," *J. Personality and Social Psychology,* vol. 64, pp. 552-560, 1993.
[8] P. Goodwin, "Enhancing Judgmental Sales Forecasting: The Role of Laboratory Research," *Forecasting with Judgment,* G. Wright and P. Goodwin, eds., pp. 91-112, John Wiley & Sons, 1998.
[9] H.P. Grice, "Logic and Conversation," *Syntax and Semantics 3: Speech Acts,* P. Cole and J.L. Morgan, eds., pp. 41-58, Academic Press, 1975.
[10] S. Grimstad and M. Jørgensen, "The Impact of Irrelevant Information on Estimates of Software Development Effort," *Proc. Australian Software Eng. Conf.,* 2007.
[11] F.J. Heemstra and R.J. Kusters, "Function Point Analysis: Evaluation of a Software Cost Estimation Model," *European J. Information Systems,* vol. 1, no. 4, pp. 223-237, 1991.
[12] C.E Hewitt, N. Mitchell, and D.J. Torgerson, "Listen to the Data When Results Are Not Significant," *British Medical J.,* vol. 336, no. 7634, pp. 23-25, 2008.
[13] J. Hihn and H. Habib-Agahi, "Cost Estimation of Software Intensive Projects: A Survey of Current Practices," *Proc. Int'l Conf. Software Eng.,* 1991.
[14] M. Jørgensen, "A Review of Studies on Expert Estimation of Software Development Effort," *J. Systems and Software,* vol. 70, nos. 1/2, pp. 37-60, 2004.
[15] M. Jørgensen, "The 'Magic Step' of Judgment-Based Software Effort Estimation," *Proc. Int'l Conf. Cognitive Economics,* 2005.
[16] M. Jørgensen and G. Carelius, "An Empirical Study of Software Project Bidding," *IEEE Trans. Software Eng.,* vol. 30, no. 12, pp. 953-969, Dec. 2004.
[17] M. Jørgensen and S. Grimstad, "Avoiding Irrelevant and Misleading Information when Estimating Development Effort," *IEEE Software,* vol. 25, no. 3, pp. 78-83, May/June 2008.
[18] M. Jørgensen and S. Grimstad, "Judgment-Updating among Software Professionals," *Proc. Second Int'l Conf. Software Knowledge Information Management and Applications,* 2008.
[19] M. Jørgensen and A.B. Kanten, "Temporal Distance and Software Development Effort Estimation," 2009.
[20] M. Jørgensen and D.I.K. Sjøberg, "Impact of Effort Estimates on Software Project Work," *Information and Software Technology,* vol. 43, no. 15, pp. 939-948, 2001.
[21] M. Jørgensen and D.I.K. Sjøberg, "The Impact of Customer Expectation on Software Development Effort Estimates," *Int'l J. Project Management,* vol. 22, pp. 317-325, 2004.
[22] D.J. Koehler, "Explanation, Imagination, and Confidence in Judgment," *Psychological Bull.,* vol. 110, no. 3, pp. 499-519, 1991.
[23] A. Liberman, M. Sagristano, and Y. Trope, "The Effect of Temporal Distance on Level of Mental Construal," *J. Experimental Social Psychology,* vol. 38, pp. 523-535, 2002.
[24] N. Liberman and Y. Trope, "The Role of Feasibility and Desirability Considerations in Near and Distant Future Decisions: A Test of Temporal Construal Theory," *J. Personality and Social Psychology,* vol. 75, pp. 5-18, 1998.
[25] T. Mussweiler, "Comparison Processes in Social Judgment: Mechanisms and Consequences," *Psychological Rev.,* vol. 110, no. 3, pp. 472-489, 2003.
[26] T. Mussweiler and F. Strack, "The Use of Category and Exemplar Knowledge in the Solution of Anchoring Tasks," *J. Personality and Social Psychology,* vol. 78, pp. 1038-1052, 2000.
[27] G.B. Northcraft and M.A. Neale, "Expert, Amateurs, and Real Estate: An Anchoring-and-Adjustment Perspective on Property Pricing Decisions," *Organizational Behavior and Human Decision Processes,* vol. 39, pp. 228-241, 1987.
[28] L.H. Putnam, "A General Empirical Solution to the Macro Software Sizing and Estimating Problem," *IEEE Trans. Software Eng.,* vol. 4, no. 4, pp. 345-361, July 1978.
[29] N. Schwarz, "In Social Context: Biases, Shortcomings, and the Logic of Conversation," *Advances in Experimental Social Psychology,* M.P. Zanna, ed., pp. 125-162, Academic Press, 1994.
[30] Y. Trope and A. Liberman, "Social Hypothesis Testing: Cognitive and Motivational Factors," *Social Psychology: Handbook of Basic Principles,* E.T. Higgins and A.W. Kurglanski, eds., pp. 239-270, Guilford Press, 1996.
[31] Y. Trope and A. Liberman, "Temporal Construal," *Psychological Rev.,* vol. 110, pp. 403-421, 2003.

Magne Jørgensen received the Diplom Ingeneur degree in Wirtschaftswissenschaften from the University of Karlsruhe, Germany, in 1988, and the Dr Scient degree in informatics from the University of Oslo, Norway, in 1994. He has 10 years' industry experience as a software developer, project leader, and manager. He is now a professor of software engineering at the University of Oslo and a member of the Software Engineering Research Group, Simula Research Laboratory, Oslo, Norway. He has published many articles on software cost estimation in international journals. He has supported work on the improvement of software project estimation and has been responsible for estimation training in several software companies.

Stein Grimstad received the MSc degree in informatics in 1998 from the University of Oslo, Norway, where he is currently working toward the PhD degree in software engineering. He has seven years' industry experience as software developer and solution architect. He is a member of the Software Engineering Research Group, Simula Research Laboratory, Oslo, Norway. His research interests include software engineering and, in particular, software cost estimation.