How to Recruit Professionals as Subjects in Software Engineering Experiments

Hans Christian Benestad, Erik Arisholm and Dag Sjøberg
Simula Research Laboratory, Department of Software Engineering
benestad@simula.no, erika@simula.no, dagsj@simula.no

Abstract. Controlled experiments are the classical scientific method for identifying cause-effect relationships, and are complementary to case studies and surveys as a means to empirically evaluate information systems development methods and practices. Most controlled experiments that evaluate software development methods and practices use students as subjects. Using students as subjects is convenient. However, a common criticism of these experiments is that it is difficult to generalize the results to industrial settings. Consequently, Simula Research Laboratory wanted to include professionals as subjects in the experiments. At present, more than 750 professional software developers from 46 software development organizations have participated in our different experiments. From this experience we have identified three important principles for research groups that want to include professional software developers as subjects in controlled experiments. First, practical constraints must be considered when defining the target population of software developers. Second, the participating organizations must be offered flexibility and value using a planned communication strategy, in order to ensure adequately sized representative samples of organizations and individuals. Third, to ensure long-term, relationships with the organizations, high professional and ethical standards must be employed. These findings complement existing knowledge on how to conduct experiments in our field.

Introduction

To support the progress of information systems development (ISD) methodologies and practices, we need to evaluate empirically their advantages and disadvantages. Case studies and surveys have been used extensively for this purpose. A third important category of empirical study is controlled experiments, the conducting of which is the classical scientific method for identifying cause-effect relationships. In controlled experiments, individuals or teams (the experimental units) conduct one or more software development tasks for the sake of comparing different populations, processes, methods, techniques, languages or tools.

At Simula Research Laboratory, we have been concerned with increasing the realism of such experiments to make the results relevant to industrial practices (Sjøberg *et al*, 2002). An important part of this is to include professional systems developers in the experiments. This is a groundbreaking approach: A survey performed by our research group showed that in the 113 experiments investigated, 87% of the subjects were students (Sjøberg *et al*, 2005). In general, attempting to generalize results from student experiments to industrial settings remains guesswork at best, since the sample is not drawn from the target population of professional software developers.

The lack of professionals in software development experiments may be due to the conception of high costs and large logistical effort. Warren Harrison (2000) puts it this way:

> Professional programmers are hard to come by and are very expensive. Thus, any study that uses more than a few professional programmers must be very well funded. Further, it is difficult to come by an adequate pool of professional developers in locations that do not have a significant software development industrial base. Even if we can somehow gather a sufficiently large group of professionals, the logistics of organizing the group into a set of experimental subjects can be daunting due to schedule and location issues.

We have met these and other challenges related to using professionals in our experiments. The main focus in this paper concerns practical issues, such as how to identify and approach companies and procedures for informing them about the experiment. However, many of these guidelines crosscut more fundamental problems related to how target populations are defined and how one samples from it.

An introduction to important aspects of experimentation in our field is given by Wohlin *et al* (2000). They visualize the experiment process as show in Figure 1.

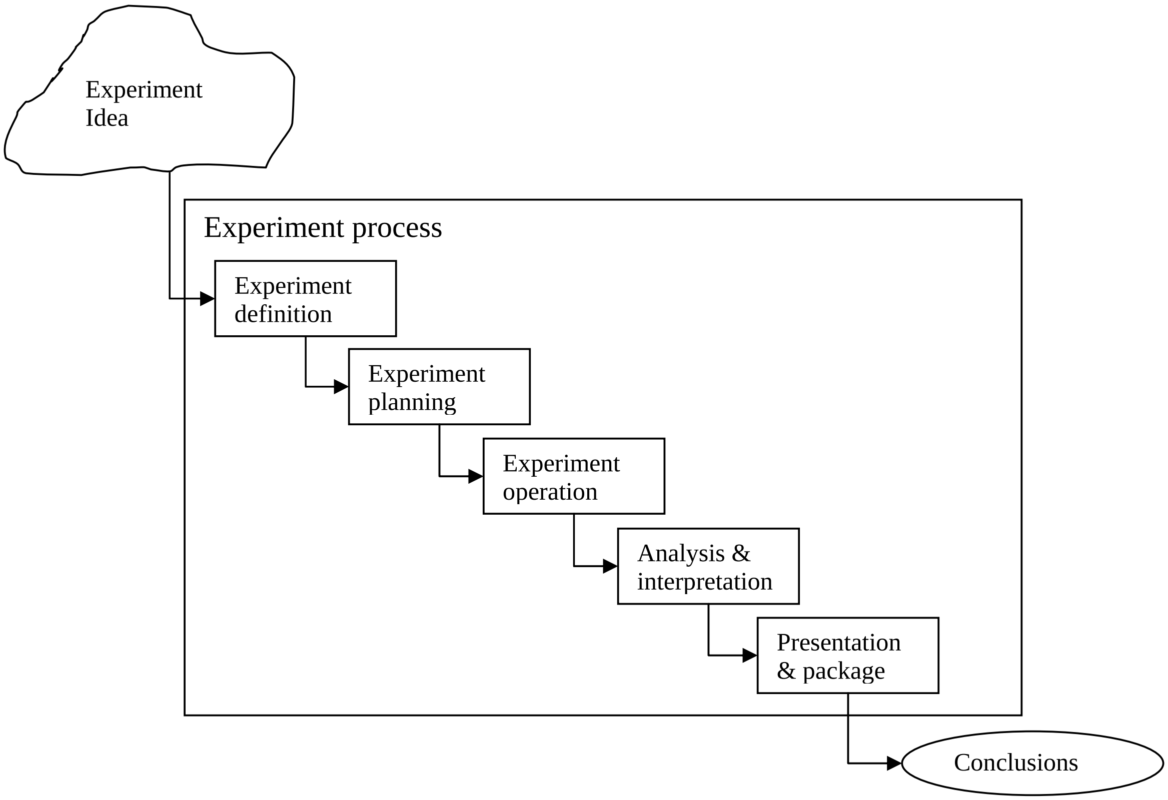

Figure 1. Overview of the experiment process (Wohlin *et al*, 2000).

The model is not a strict waterfall model. Activities may start before others are finished, and iterations may occur. During *Experiment definition* we define the experiment in terms of objective and goals. In the *Experiment planning* phase, the design of the experiment is determined, the instrumentation is considered, and the threats to the experiment are evaluated. In the *Experiment operation* phase, practical preparations for the experiments are made, then the experiment is executed and data collected. These data are analyzed and evaluated in *Analysis & interpretation*. During *Presentation & package*, information about the experiment is packaged e.g. for the purpose of making the results known to the research community or the industry, or for the purpose of replication of the experiment.

In the remainder of this paper we describe how this process is affected by our strategy of using professional software developers in the experiments. First, we describe the practical issues that must be considered when defining the target population. Then, we describe how we go about to recruit subjects in order to supply the experiment with an adequately sized representative sample from that population. Finally, we describe how we try to build long term relationships with the industry by pursuing high professional and ethical standards. Key data from selected experiments is provided at the end of the paper.

Defining the Target Population

The target population is the set of individuals, projects or organizations to which we want to generalize the results of an experiment. Definition of the target population begins during *Experiment definition* and is concluded during *Experiment planning*. If one applies statistical hypothesis testing, an underlying assumption is to have a well-defined population from which a random sample is drawn. Kitchenham (2002) states: *“If you cannot define the population from which your subject/objects are drawn, it is not possible to draw any inferences from the results of your experiment”*.

The definition of the target population guides which organizations and developers to approach in the recruiting (i.e., sampling) process. Consequently, we define it prior to this process. The availability of subjects may be constrained by practical considerations, so such constraints should be considered alongside the scientific considerations when defining the target population. In other words, the target population must be narrowed down to the set of individuals that we have a reasonable chance of recruiting in an experiment. The following sections describe these constraints based on our experience, roughly categorized into *geographic area, organizational profile* and *individual profile*.

Geographic Area

Sampling without constraints on geographic area is yet to be seen within software engineering experiments. It would require an unreasonable amount of resources to be able to randomly select software developers without such constraints. In practice, it will be necessary to limit the geographic area that we target.

Cultural differences between geographic areas will then make it hard to generalize our findings to other areas, because the differences may influence on the effect of technologies or methods that we want to study. If the ambition of the research project is to generalize results to a larger area than the proximity of the research group, there may be no other way than to include other cultural and geographical areas in the experiment. We have found the challenges in doing so to be manageable.

For instance, in one experiment, we wanted to investigate the effect of pair programming as compared to traditional individual programming. Since we could not rule out the possibility of cultural differences, we defined sub-populations from three countries (Norway, Sweden and the UK). The organizations were identified, approached and handled in very much the same manner, described later in this paper. Since researchers had to be present at the locations of the experiment, the cost of each experiment session increased by about 20%. The subjects were exposed to experiment material in their native language. This implied an extra effort of about 100 man-hours per language. This extra cost and effort were considered to be reasonable from a cost-benefit perspective. More distant countries or cultural areas were not included in this experiment, but plans exist for replicating the experiment in typical outsourcing countries like India. The project will be able to analyze potential differences between the countries. If it turns out that the results do not differ significantly between the countries, it can be used as an indication that cultural differences are less important for the research question at hand.

Organizational Profile

The organizational profile has an impact on how software is developed. For example, in-house development of off-the-shelf software may differ from contract-based development of customer specific solutions. Creating embedded software for the space industry is different from creating online services for the bank industry. Large, delocalized companies might have different work processes than have they in small, local companies.

The impact of such differences may depend on the research questions at hand. In many situations, it may be both necessary and sensible to limit the kind of organizations that participate in the study. By selecting subjects from homogenous organizations, results can be more reliable for that kind of organizations. Also, the research questions may be of particular relevance to some kinds of software organizations, hence motivating constraints on organizational profiles included in the target population.

For example, in some experiments we have limited the target population to consultancy companies developing customer specific solutions. This kind of organizations constitutes a major part of the Norwegian IT-industry, so our studies still target an important industry sector. The organizations find our studies relevant, because we target their kind of organizations.

Individual Profile

Many methods, techniques or tools that we want to study, are likely to have different effect on people with different skills, experience or other characteristics. For example, some techniques can be expected to increase productivity for a junior developer, while some other technique may be beneficial only for the advanced developer (Arisholm and Sjøberg, 2004).

The problem of identifying which, if any, personal characteristics to include in the definition of the target population, or how to define sub-populations based on them, is beyond the scope of this paper. Still, the practical constraint of whether people with the defined skills are likely to be available (within the geographic area and kinds of organizations targeted), must be considered. Also, if personal characteristics are part of the definitions of the target population, we must be able to assess every candidate on these characteristics. This is described further in section “Selecting Subjects Within Organizations”.

Recruiting Participants

Given a description of the target population from *Experiment planning*, the first part of *Experiment operation* is concerned with obtaining commitment for participation from representatives of this population. In controlled experiments we typically study a sample from a larger population and use inferential statistics to make an inference from the sample back to the population. Kitchenham (2002) states that “The subjects and objects must be representative of the population or you cannot draw any conclusions from the results of your experiment”. This ideal model is not simple to adhere to in practice. Professional software developers are not immediately available for random sampling to the research project! So in practice, we have followed a two-step selection process: First, we identify and select organizations to approach. Then we select subjects from within each organization.

Identifying Candidate Organizations

Once the target population has been defined, candidate organizations can be identified. Ideally, we should identify organizations that in sum employ the total target population. It may be perceived easier to contact and recruit from organizations known by the research group. This network can be used, but only as one of several sources for identifying potential participating organizations. Approaching organizations purely based on acquaintances may introduce an undesired and unnecessary bias. We have not observed a significant difference in participation rate (number of organizations that agreed to participate, divided by the number of organizations invited) between organizations and people we knew in advance, and the ones we did not know.

Information available on the internet is our main source for systematic identification of relevant organizations. We can expect every organization engaging in software development to share information on its business on the internet. Finding them is a matter of using general internet search, or using specialized databases of vendors, associations, partners and interest groups. Investigating internet sites of candidates will usually give a fair indication of whether their profile matches the defined target population of the experiment.

We systematically track information on contacts we make with the organizations. Reusing contacts from earlier experiment collaboration can be effective. Doing this, it is important to adhere to the principles of unbiased selection described above: Every company identified should have an equal chance to be selected for a given experiment.

Marketing the Project

When asking professional developers to participate in software development experiments, we are competing for access to scarce resources. Because of this, there is always a risk of introducing a bias in the selection of participants. For instance, less-successful companies may have more time available to participate. Alternatively, technology innovators or organizations already engaging in research activities may be more interested in participating. Those companies, and their employees, may not be completely representative of the defined target population.

We attempt to address this selection bias by increasing the attractiveness of the experiments. Our strategy for achieving this is by offering *flexibility* and *value* using a planned *communication strategy*. The actual offer can be tailored to the specific organization and situation, based on our knowledge of its business, and general market conditions.

Flexibility

From a given organization we will require some number of individuals, with some profile, at some time and location and for some duration. Participation will be more attractive if we can be flexible on these variables. Some of the variables may be fixed by experimental design, notably individual profile and duration of the experiment. For most of our experiments, however, we have been able to let the participants choose among several dates, within a time period of four to six months. This has been one of the best selling points for our projects. The organizations were able to plan for participation well in advance, utilizing temporary demand dips.

By using an internet-based experiment tool, we are able to offer flexibility when it comes to location of the experiment. Time and cost are reduced by executing the experiment at the premises of the participating organizations. The internet-based *Simula Experiment Support Environment (SESE)* and experiences with its use is described by Arisholm *et al* (2002). To ensure control, we are still physically present at the site of the experiment.

A project may state initial ambitions on maximum or minimum number of participants from each organization. If such numbers are based on practical considerations, we treat the limits flexibly in order to achieve our overall goals.

Value

In order to obtain consistent and broad participation from the industry, the experiments must give something of value back to the participating organizations. “What’s in it for us?” is a question we must be well prepared to answer. A perceived acceptable answer largely depends on the context. For some organizations, commercial value would be most important, for others the acquisition of new knowledge and skills would be more attractive.

In many experiments, we have offered payment related to the total number of hours or days spent on the experiment. This has been a successful way of recruiting organizations and employees normally doing time-and-material assignments. For such organizations, experiment participation can be covered by existing business processes for handling leads, writing contracts, allocating resources, and billing.

For other kinds of businesses, for instance for those doing in-house development, or for those fulfilling large fixed-price customer contracts, direct payment does not necessarily make the experiment sufficiently attractive. The opportunity of acquiring new knowledge can be used as an additional selling point. In some experiments, the participants are likely to gain new skills from the sole participation. For example, in one experiment we introduced the developers to the concepts of pair programming, and let them practice for one day. We believe the effect of this was close to a targeted one-day course in pair programming.

In most projects we promise the participating organizations to gain early access to research results from the experiments. If the research questions are sufficiently relevant, this can be attractive. When research results are ready for presentation, we organize free seminars exclusively targeting the participating organizations. Many organizations appreciate this, and it gives us an extra opportunity for building long-term relationships.

A third variant is to utilize resources in the research group, to offer free courses, seminars or mentorship on themes of interest to the organizations. This can also help building long term relationships with participants.

We recognize the fact that most research institutions in our research area do not have the necessary funding for paying professionals participants in experiments, so the alternative approaches may be more affordable in the short term. However, given the importance and challenges of the information system industry, we believe more research groups should apply national and multi-national (e.g., EU) research bodies to fund realistic experiments.

Communication Strategy

Having identified our needs and possible offerings, these must be communicated in a well-planned manner. An efficient communication strategy specifies how contacts are to be made, how commitment is ensured, and how information on the experiment is given.

How contacts are made: Based on the list of candidate organizations, enquires for participation can be made. The purpose of the first contact is to create a positive interest and arouse curiosity. This is planned by answering the following questions:

- Who makes the first enquiry?
- Who is the first enquiry made to?

- What medium is used for the first enquiry?
- Who makes the first enquiry?

The answers depend on the situation at hand, and there are also interdependencies between the answers. A summary of our experiences and considerations related to these questions are given below.

Letting the enquiry come from a research director may have greater effect than letting it come from a research assistant.

We have found that it works well to make enquiries to the higher levels of company hierarchies, identifying people that control resources of interest to the project, and that normally handle external relations. Using the switchboard and asking questions like “Who manages java resources in your company” has been successful. For smaller companies (10-50 employees) we have normally contacted the CEO. If the contact point belongs to a low level of the company, there is a danger of selecting from a specific sub-culture, or the person may not have necessary incentive or power to attract potential participants.

For some projects we have initiated contact by phone, for others we have used email. It may be safer to initiate contact by email, because it can be perceived as less intrusive, and the wording can be well planned. However, for people with good communicative qualities, phone calls can be effective and efficient. Whatever method chosen, the mindset of the sender should resemble that of a salesman more than that of a researcher.

As to the form and content of the first message, we try to make it short and concise. The following line contains enough information for a first contact (taken from one mail correspondence):

> ”We would like to hire between 6 and 24 java developers at one agreed upon date and rate, to participate in a large scale international research project this autumn. Please contact …”

If the enquiry is expressed so that the receiver classifies it as a “lead”, it will normally be handled by a well-defined business process. If we do not receive a reply on an initial enquiry made by email, we follow up by phone.

How commitment is ensured: The goal of the first enquiry is to create positive interest in the experiment. After this, interested organizations request more information, or will have specific questions related to the research project. Once such feedback is received, the parties will have a common interest in concluding whether the requirements from the research group can be matched, and whether the value received is acceptable.

We generally find oral communication more efficient in this phase, and it also improves the chance of establishing a good personal relation. The concrete set of requirements, and measures to make the experiments attractive, must be considered and communicated. It may also be necessary to consider special adaptations for each individual organization.

How information is given: The essential information about the experiment is most often given while making the first contact. Later on, information related to preparations and the practical experiment execution is given. At this point, a coordinator in each company can be given the responsibilities of distributing information and ensuring that necessary preparations are carried out. This person may be different from the initial contact. The responsibilities of the coordinator can be part of a contract if they are critical to the project. To further ensure that enough resources are put in the preparations, time used by the coordinator can be made billable.

To a great extent, the information given to the companies is prepared well in advance, during the experiment design phase of the research project. This is both for efficiency reasons and because information given to participants constitutes an important part of the experiment. Examples of materials that can be prepared are:

- An email template for the first contact
- A contract template
- General information on the experiment, with requirements to the participants
- Information/checklist to the coordinator
- Information/checklist to the participants

The material can be subject to refinements as feedback and experience is collected. However, every change is considered from the perspective of whether it will introduce an undesired bias among the participants, or any other methodological problem. Misunderstandings are good feedback, in that they clearly signal needs for improvements when it comes to information. For example, in a recent project, the coordinator wanted to send only their top developers, because they "didn't want to be put in bad light". Obviously, more information on anonymity was needed.

Selecting Subjects Within Organizations

Once commitment is achieved, there are several challenges related to the selection of subjects from within each organization. Ideally, we should recruit subjects from the organizations in such a way that our total sample of subjects is representative of our predefined target population requirements.

If personal characteristics are part of the definition of the target population, or used as criterion for dividing into sub-populations (e.g., different levels of seniority), the first challenge would be to assess every candidate on these characteristics to determine whether the candidate meets our requirements.

Sometimes, it is possible to define the target population using quantitative criteria. "Years of programming experience", "java certified" or "years of higher education" can normally be read from the resumes of the candidates. However, we know that many individual differences related to skills are not well captured by such measures. Assessment of skills is non-trivial. The fact that assessment must be done ahead of the experiment makes it even more challenging.

In one experiment, we defined sub-populations of junior, intermediate and senior Java consultants. The judgment made by their closest manager was used for the categorization. Empirical data from the experiment itself showed that this judgment was meaningful. In general, when the assessment cannot be performed by straightforward quantification, we expect such judgment to be a cost efficient and reasonably reliable way of assessing the characteristics of the candidates.

Although we know who is included in the target population, we meet the challenge of *availability*. Highly valued employees may not easily be released from their normal work, while the low-skilled or inexperienced ones may be available. Alternatively, high-skilled developers closely following research can be more interested in participating. Both these scenarios may introduce bias in the selection.

The already described recruiting strategies of *flexibility, value* and *informing* address this challenge as well. If participation in the experiment is perceived as more attractive, the organizations will be more willing to assist in selecting a more representative sample, and the individuals themselves will be more willing to participate.

Building Long-Term Relationships

The previous section described how we recruit subjects for one given project. The way we approach and handle organizations in one project will have an impact on the responses on later occasions. By the time of writing, more than 750 professional software developers have participated in our experiments, representing 46 software development organizations. In one single experiment, we approached 62 organizations. Both the 42 that did not participate on this occasion, and the 20 that did, may be targets for future experiments. The way we manage these relations will strongly influence their future attitude towards us. By pursuing high ethical standards and levels of professionalism in all phases of the experiment, we try to build and maintain long-term relationships. We elaborate further on these issues below.

Building Public Image and Awareness

A well-reputed research group will undoubtedly benefit from its reputation when approaching organizations for participation in software development experiments. The public image of a research group engaging in empirical research in industry is one of its most valuable assets. Building it goes further than through the studies themselves. The way the group appears at its own website, in scientific papers, in teaching activities, at seminars, in industry collaboration projects, and in experiments all shape the public view of the group.

Ethical Concerns

Disregarding ethical concerns will sooner or later make recruiting to software development experiments very difficult. Ethics in controlled experiments with industrial participants have been discussed in a broader context by Singer and Vinson (2002), and Moløkken-Østvold (2005). Perhaps the primary ethical principle in human research is that of *full informed consent*: Participation should be a free decision based on information describing any effect participation may have on the subjects. Elaborate investigation is needed to determine the effects an experiment may have on the subjects. Once discovered, we might be able to take measures to remove or reduce these effects. For example, when solving tasks in an experiment, subjects may discover that colleagues use more or less time than themselves. The perceived ability of the subjects to draw such conclusions should be reduced, for instance by informing them specifically that all subjects do not solve identical tasks. Ensuring that the proposed information does not influence the outcome of the experiment in any unforeseen way also requires elaborate investigation. Conflict arises when removal of some part of the information violates the principle of informed consent.

These are all difficult considerations, increasing the need for the research group to plan the information strategy well in advance. A further complicating factor is that there is normally two levels of information, one for the employer and one for the employee.

The topic of voluntariness is also challenging. For example, participants may become under pressure from their employer to participate, or to avoid terminating their participation. It is the responsibility of the research group to demand voluntariness, and agree with the employer that participants may discontinue the experiment at any time.

A third topic is related to a qualitative assessment of individual skills, which was discussed earlier in this paper. Such assessment can be sensitive, and we must take measures to avoid discontentment: Developers will not find it pleasing to be put in a group of “low-skilled developers”. These things must be discussed between the research group and the employer, since the solution may depend on the organizational culture. It is not unusual that organizations are asked for people with different levels of skill, but we realize that a research project can be perceived differently from an “ordinary” software development project.

Executing the Experiment

Execution of the experiment is part of the *Experiment operation* phase from Figure 1. It can be a challenging task. Among the things to consider are planning the reception and departure, setting up and testing the technical environment, setting up work places, ensuring appropriate staffing, and organizing breaks. Writing down guidelines and maintaining checklists improves quality, saves time, improves the long-term relationships with the organizations, and ultimately, the research.

Taking part in a controlled experiment can be very different from the normal working situation of the subjects. A welcome session is used to make the subjects feel comfortable, and to ensure that they are adequately informed. We stress the importance of their presence, and acknowledge their willingness to participate. Confidentiality issues are stressed, and this is also a good time for repeating the opportunity of termination.

Our research group has developed the experiment tool, SESE (Arisholm *et al.*, 2002) which has been essential in the conduct of large, geographically spread controlled experiments. The tool provides the subjects with all experiment materials and all data is collected in a controlled manner, over the internet. The research group can monitor the progress of every individual. With the tool, it is possible to define the logistics of the experiment in advance, saving valuable time in the experiment. Minimizing idle-time or wasted time in experiments make a good impression on the participants. It also ensures that the subjects follow the required procedures, which in turn improves the quality of the collected data.

Wrapping Up an Experiment

Having executed the experiment, data should be validated. The validation deals with aspects such as if the participants have understood the tasks, and that they seem to have participated seriously in the experiment (Wohlin *et al*, 2000). A successful execution of an experiment is a major achievement. Summarizing experiences and loose ends is still important:

- Give feedback and ask for feedback
- Summarize contractual obligations that should be fulfilled
- Leave an open end for further collaboration

These activities conclude the *Experiment operation* phase from Figure 1.

Giving positive feedback to participants is important as a means to build and maintain long-term relationships. Some participants may feel that they did not perform as well as they should in specific tasks. Feedback should stress that their willingness to participate, and their ability to follow the experiment process, is what was important to the experiment. Negative feedback should have a planned, rational purpose. For example, if half of the experiment participants did not show up, a purpose could be to agree on a future experiment day.

Asking for feedback both has the purpose of being a source of suggestions for improvements, and of showing care and interest in the viewpoints of the participants. Some kinds of feedback may also need immediate attention, for instance if the purpose of the experiment is clearly misunderstood.

The research group may have specific contractual obligations to fulfill after the experiment. Whether these are related to payment, delivery of services or sharing of research results, they should be as clearly stated as possible. Sharing of results will normally belong to the *Presentation & package* phase from Figure 1. Obligations from the participants may include non-disclosure agreements that have been made. The essence of these can be repeated at this stage.

Finally, leaving an open end for future collaboration is helpful for future projects. It is also a good way of acknowledging that the research group was happy with the collaboration.

Supporting Organization and Tools

The focus on using professionals in our experiments has largely influenced how we organize our research group and research projects. We have employed an experienced project manager (the main author), having as one of his responsibilities to administratively manage and to recruit to our research projects. This ensures cross-project learning, and a long-term focus on building relationships with the industry. It also helps in balancing research consideration with practical constraints when designing experiments.

We find the SESE-tool invaluable for the planning, execution and analyzes of the experiment. For the recruiting process itself we evaluated CRM tools to manage industry-relations, concluding that they were oversized for our needs. E-mail and phone are the basic tools employed. For the most challenging projects we use a simple web-based issue tracker to be able to manage the status and contact history of each company relation. After each project we extract lessons learned in a simple web-based knowledge management tool. The information resident in those tools is of great benefit for new projects, and is also the main source for the quantitative data given in the next section.

Example Projects with Key Data

Table I summarizes key data from five experiments using professional software developers as subjects. They are increasingly complex with regards to the task of recruiting professional subjects.

"BESTPro" is an example from a series of studies investigating theories related to effort estimation. It is a special case, in that it involved only one company. It is further described by Jørgensen (2004).

"Design Patterns" investigated whether the presence of documented design patterns combined with the subjects' knowledge of design patterns had any effect on the time and quality of maintenance tasks. The experiment is further described by Vokac *et al* (2004).

"BESTWeb", an experiment investigating the effects of UML-documentation, targeted medium-sized and large consultancies in Norway. Up to three developers from each organization were selected. Results from this experiment are not published at the time of writing.

"Ind", an experiment investigating the effects of different control styles in java programs, targeted medium-sized and large consultancies in Norway and Sweden, and included every developer qualified according to the inclusion criteria. The experiment is further described by Arisholm and Sjøberg (2004).

"PPE", an experiment investigating the effects of pair programming, targeted medium-sized to large consultancies in Norway, Sweden and the UK and included every developer qualified according to the inclusion criteria. Results from this experiment are not published at the time of writing.

Table I, Summary of selected controlled experiments at Simula Research Laboratory

	BESTPro	Design Patterns	BESTWeb	Ind	PPE
Study period	February 2002	13-15 May, 2002	January-June 2005	November 2001-February 2002	May 2004-January 2005
Study duration (per participant)	3 hours	3 days	5-8 days	5-8 hours	5-8 hours
Commercial terms	€80/hour	€70-90/hour	€100/hour	€70-90/hour	€70-90/hour
Study locations	Site of participant	Simula	Simula	Site of participants	¾ site of participants ¼ Simula
Total number of professionals	16	44	20	124	180
Individual profile	7 project managers 9 senior software developers	C++ programmers	Senior Java consultants Struts/JSP Eclipse UML Mysql	Java consultants, 1/3 junior 1/3 intermediate 1/3 senior	Java consultants, 1/3 junior 1/3 intermediate 1/3 senior
Organizational profile	Consultancy	Consultancies + Research institution	Consultancies 10-20 companies	Consultancies	Consultancies
Geographic area	Oslo	Oslo	Oslo	6/7 Norway 1/7 Sweden	4/10 Norway 3/10 Sweden 3/10 UK
Recruiting period	December 2001	December 2001-April 2002	November 2004-January 2005	September-December 2001	March -October 2004
Organizational effort	1 day	2 weeks	5 days	12 days	25 days

Conclusions and Further Work

We have found the key challenges related to using professionals in software development experiments to come in the categories of

- Balancing scientific ambitions with practical constraints when designing the experiment
- Making the experiments sufficiently attractive to the software industry
- Building long term relationships with the industry by pursuing high professional and ethical standards

We believe that these challenges are essential to research groups with the ambition of including professionals in software development experiments. The way we meet the challenges is not necessarily optimal for any research group, but can be used as a source of inspiration and ideas.

Over the past years we have moved from achievements by individuals, to a well-planned, repeatable process. In the future we will work further in this direction, with the goal of increasing the ratio between the scientific value, and the cost and effort implied.

References

Arisholm, E., Sjøberg, D., Carelius, G., and Lindsjørn, Y. (2002). "SESE – an Experiment Support Environment for Evaluating Software Engineering Technologies", *NWPER'2002 (Tenth Nordic Workshop on Programming and Software Development Tools and Techniques)*, pp. 81-98, Copenhagen, Denmark, 18-20 August, 2002.

Arisholm, E. and Sjøberg, D. (2004). "Evaluating the Effect of a Delegated versus centralized Control Style on the Maintainability of Object-Oriented Software", *IEEE Transactions on Software Engineering*, vol. 30, no. 8, 521-534.

Harrison, W. (2000). "N=1: An Alternative for Software Engineering Research?, Beg, Borrow or Steal: Using Multidisciplinary Approaches in Empirical Software Engineering Research", *Workshop, 5 June, 2000 at 22nd International Conference on Software Engineering (ICSE)*, Limerick, Ireland.

Jørgensen, M. (2004). "Top-Down and Bottom-Up Expert Estimation of Software Development Effort", *Journal of Information and Software Technology*, vol. 46, no. 1, 3-16

Kitchenham, B.A., Pfleeger, S.L., Pickard, L.M., Jones, P.W., Hoaglin, D.C., El Emam, K., and Rosenberg, J. (2002). "Preliminary guidelines for empirical research in software engineering", *IEEE Transactions on Software Engineering*, vol. 28, no. 8, 721-734.

Moløkken-Østvold, K. (2005). "Ethical Concerns when Increasing Realism in Controlled Experiments with Industrial Participants", *Proceedings of the 38th Hawaii International Conference on System Sciences*

Singer, J. and Vinson, N.G. (2002). "Ethical Issues in Empirical Studies of Software Engineering", *IEEE Transactions on Software Engineering*, vol. 28, no. 12, 1171-1180.

Sjøberg D., Anda B., Arisholm, E., Dybå, T., Jørgensen, M., Karahasanovic, A., Koren, E., and Vokác, M. (2002). "Conducting Realistic Experiments in Software Engineering", *proc. ISESE'2002 (First International Symposium on Empirical Software Engineering),* October 3-4, 2002, pp. 17-26.

Sjøberg, D., Hannay J., Hansen, O., By Kampenes, V., Karahasanovi¢, A., Liborg, N., and Rekdal, A.C. (2005). "A Survey of Controlled Experiments in Software Engineering", Submitted to *IEEE Transactions on Software Engineering.*

Vokác, M., Tichy, W., Sjøberg, D., Arisholm, E., and Aldrin, M. (2004). "A Controlled Experiment Comparing the Maintainability of Programs Designed with and without Design Patterns – a Replication in a real Programming Environment." *Empirical Software Engineering,* vol. 9, no. 3, 149-195.

Wohlin, C., Runeson, P., Höst M., Ohlsson, M. C., Regnell, B., and Wesslén, A. (2000). *"Experimentation in Software Engineering: An Introduction."* Kluwer Academic Publishers.