A Software Engineering Experiments to value MDE in testing. Learning Lessons

M.J. Escalona[(1)], G. Lopez[(1,2)], S.Vegas[(3)], L. García-Borgoñon[(1,2)], J. A. García-García[(1)], N. Juristo[(3)]

[(1)]Grupo IWT2. Universidad de Sevilla
[(2)]Instituto Tecnológico de Aragón
[(3)]Universidad Politécnica de Madrid
mjescalona@us.es, glopez@itainnova.es, svega@upm.es, laurag@itannova.es, julian.garcia@iwt2.org, juristo@upm.es

ABSTRACT
Controlled experiments are commonly used to evaluate Software Engineering methods, processes and tools. Validating results of Software Engineering research in industrial settings is a direct way to obtain feedback about its value. However, few software engineering experiments are running in industry. The lack of communication between companies and research teams does not make the necessary cooperation among them possible. This paper presents our experiences when running an experiment dealing with Early Testing at the University of Seville. It also introduces the strategy we followed to obtain the participation of 97 practitioners from 32 different software companies. Such strategy is pointed out as a set of guidelines to successfully involve this large number of companies and practitioners.

Keywords: Software experiments in enterprise, early testing experiences

1 INTRODUCTION

Empirical studies are becoming increasingly common in Software Engineering and the acceptance that their contributions are enriching the knowledge of this area is continuously increasing. Empirical studies are needed to develop or improve processes, methods and tools for software development and maintenance [1].

Controlled experiment is a type of empirical study. It allows the identification of cause-effect relationships [24]. The experimental paradigm proposes that laboratory findings should be generalized through other types of experiments closer to real world. In Software Engineering, the equivalent to field experiments is experimentation in industry [26].

A major ambition of experimental Software Engineering is to provide software managers, who are in charge of decision-making processes in software development industry, with evidence on how new technology can be introduced [16]. However, the experimental subjects of these experiments are typically studied with little or no professional experience [3]. The few experiments carried out with practitioners are isolated experiences [26]. Additionally, the experimental setting and materials tend to be artificial or only partially related to real projects [24].

The systematic use of experiments as a way to face the decision analysis and resolution practices is far from what would be desirable.

In [26], authors report preliminary results regarding the difficulties encountered when they run experiments in the industry. One of their findings is that reporting used in scientific journals is not appropriate for practitioners and what is more important, frequently in reported experiences the number of participants is very low and research community has real problems to evolve a good number of them in experiments.

The goal of this paper is to present a set of guidelines (good practices) for involving a large number of experimental subjects and companies when running experiments with practitioners. To illustrate these guidelines, we present a real experiment conducted at the University of Seville, with a total of 97 practitioners registered (76 of them finally participating in the experiment) from 32 different software companies to evaluate our research results in Early Testing.

The reminder of this paper is organized as follows: Section 2 presents both, a general view of software experiment and the motivation for our experiment. Section 3 describes our experiment in detail, and Section 4 numbers and explains the mechanisms and strategies followed to get an effective participation of companies. Section 5 summarizes learned lessons and finally, the paper ends up by stating conclusions and future lines of work in Section 6. It is important to stick out that this paper does not aim to present the results of our experiment, but to recommend a strategy that can help research teams to involve a large number of practitioners and companies in software engineering experiments.

2 BACKGROUND

Dieste *et al*. [3] report the preliminary results of a systematic literature review exploring the features of experiments run in industry. They have located a total of 15 studies. The results highlight that the software engineering community perceives experimenting in industry as a problematic activity: ***few experiments have been run and companies are demotivated to carry them out since they do not realize their value and direct benefits***. Jedlitschka *et al.* [15] empirically develop a model that characterizes software engineering practitioners information needs. They evaluate the effectiveness of the proposal with 22 software managers. ***The results have shown that experiments can be a valuable source of information in industry environments, although they do not solve the problem of involving companies in conducting them.***

Vegas *et al.* [26] discuss the difficulties they identify when running experiments in software industry. One of their main finding is the difficulty to involve a large number of participants from industry. In view of these results, it seems that experiments could be valuable for the software industry. However, few experiments are run in industry, and they do have a small number of participants. Therefore, it is essential to find a way to encourage software engineering companies and practitioners to participate in experiments. We planned an experiment taking advantage of the relationship some of us have with some software development companies. The Web Engineering and Early Testing (IWT2) research group [27] has an extensive experience in technology transfer and has set a very fruitful relation with companies through their joint collaboration in a large number of software projects [6]. Thus, we conducted an experiment with a main objective: to validate our research in Early Testing. Besides, we wanted to value some strategies in order to make practitioners participate in that project. In fact, this paper mainly focuses on this secondary objective. For that purpose, we present the strategy we have followed to successfully enrol a large number of industrial participants in our experiment.

3 DESCRIPTION OF THE EXPERIMENT

3.1 Experiment Definition

The goal of the experiment is to ***analyze the adequacy of the paradigm driven by models (or MDE-Model Driven Engineering) [25] as the basic technique for developing functional tests in an early phase (Early Testing)***. The experiment was opened to all software engineers who do some kind of testing in their daily work. Initially, 97 participants registered in the project, although some of them were not able to assist. Finally, 76 practitioners from 32 different software companies participated in the experiment. The call for the experiment was opened to staff of companies (small, medium and large) and software organizations (private and public) based anywhere in Spain. Participants mainly came from Andalusia, but people from other areas like Madrid or Barcelona also took part in the project. They were not required a particular profile, apart from being a software engineer who performs some kind of software testing in their daily work. Thus, juniors and seniors testers, test managers or developers who do testing in a timely manner were able to participate. That heterogeneous group required the collection of demographic data of each participant in order to guarantee a proper randomization that gave us a right profile of each participant. For this purpose, an electronic form was used for registration[1]. The contents of this form are summarized in Table 1.

Table 1. Data gathered in the registration form.

Personal Data	• Name • Surname • NIF[2] • Date of Birth • Telephone number • Email address • Gender
Academic Data	• Degree • Date of degree • Any special course related to testing (such as TestQA)
Professional Data	• Company • Position: − Manager − Analyst − Designer − Programmer − Scholar employer − Other • Years of experience in testing or in other fields (like programming or commercial activities, among others) − Experience in testing: − Test manager

[1] The questionnaire was designed in Spanish and it was developed using Opina[19]. It is available in http://iwt2.org/opina/c/290

[2] NIF (Número de Identificación Fiscal) is a personal number that each person is assigned in Spain by the Government.

	− Junior tester − Senior tester • Percentage of time spent in testing per week: − Less than 25% − Between 25% and 50% − Between 50% and 75% − More than 75% • Typical testing performed in his/her work: • Unit testing • Integration testing • Functional testing • System integration • Acceptance testing

There are two research questions in the experiment:

- **RQ1:** Is the developer effort affected by MDE when generating functional test cases?
- **RQ2:** Is the quality of the functional test cases generated affected by MDE?

Each of these research questions has an associated hypothesis (H):

- **H01**: The effort to build a test case using MDE is the same as using manual development.
- **H02**: The control and quality of test cases generated using MDE is the same as using manual development.

The experiment includes one relevant factor: functional definition method. Control is a manual method while treatment is a MDE method supported by an NDT-Suite [10] solution. NDT-Suite consists of a set of tools to support NDT (Navigational Development Tecniques)[5][7]. It is not the aim of this paper to present NDT in detail, nevertheless, we can briefly introduce that it is a model-driven methodology that offers suitable support for the application of MDE in software development. NDT covers the whole software lifecycle, from requirements to maintenance, and it even supports other management activities like quality assurance, project management and software security assurance. NDT has been successful applied to a large number of real projects, even though its suitability for companies was never methodologically valued at companies. As NDT covers many tools and software phases, we selected the most interesting ones for our first experimentation with companies. Besides, it must be added that one of the most important characteristics of NDT is the support it provides for Early Testing activities. Testing is currently a critical phase for software companies because they usually have poor resources [11]. Thus, we selected concretely the functional phase. The experiment aimed to test whether, in fact, experts consider that these solutions are suitable enough for enterprise environments.

The response variables measured in the experiment are:

- RQ1 required measuring the developer effort. The selected metric was the ***resolution time***.
- RQ2 was a little more complicated. Initially, as our experimental objects came from real projects (as explained in section 3.2.1), we had the code of the system in the first version (and also in the final one). Thus, if experts detected a good number of errors or mistakes, we considered that they defined a good set of functional tests. However, it was not possible. Although this aspect will be treated in section 4, tests defined manually by our experts did not detect any errors (the reason will be analyzed later), so we reconsidered this way of evaluation and we decided that two experts would assess the functional test definition, would value the participant tests and would give them a mark that ranged from 0 (very bad) to 4 (very good).

3.2 Experimental Design

The experiment follows one factor-two level (control and treatment) within subjects design.

Table 2 presents the experimental operation. The experiment had two differentiated parts. In the first part, participants were asked to exercise the control. This implied to generate functional tests manually. We did not offer any support or extra learning because we aimed to assess the knowledge of our experts in that field. In the second part, after a training session, we asked participants to generate functional tests using MDE with NDT. After that, each participant filled out a form.

We took into account several points before defining the schedule:

1. We had to offer, at least, two different software systems; one manual and other automatic. Thus, we first selected PIF (Information Flamenco Point)[21] and Ambassador Hotel [12] projects. Nonetheless, there was a first idea learned for that selection. When we designed our experiment, we considered that the systems should be easy to understand by our participants and they should be real systems to be more attractive for them. Thus, we were thinking about other true projects in which our group collaborated like aeronautical projects [8] or health projects [18]. Finally, we discarded them because their functionality required a previous training session on the software application domain, and we preferred to involve a big group of different practitioners. That was the right decision, as the chosen systems did not require extra time to explain our use cases to participants.
2. The complexity of the system was also another important decision. We selected a set of use cases from each of our systems; two easy use cases (one from each system) and two difficult ones (also one from each system). Thus, we had four different tasks for our experiment:
    a. Ambassador Easy, we coded as D1F
    b. Ambassador Difficult, we coded as DID
    c. PIF Easy, we coded as D2F
    d. PIF Difficult, we coded as D2D

Table 2. Timetable for the experiment[3].

Time	
15:30-16:00	Welcome and presentation
16:00-16:30	Presentation of examples and material
16:30-17:00	Defining Functional Testing manually
17:00-17:30	Coffee breack
17:30-19:00	Training course: knowing MDE. Presenting NDT and NDT-Suite
19:00-19:30	Defining Functional Testing automatically
19:30-19:45	Final interview
19:45-20:00	Closing and final conclusions

According to the experiment definition presented in Table 2, each participant must perform two tasks, one manually, and one automatically using MDE, concretely, NDT and NDT-Suite as Table 3 shows.

[3] This table presents an afternoon section. Further information can be found in the website of the experiment: http://iwt2.org/experimentacion-en-testing-temprano/.

.

Table 3. Experimental design.

	Manual	**Automatic**
Group 1	D1F	D2F
Group 2	D2F	D1F
Group 3	D1D	D2D
Group 4	D2D	D1D

Subjects were assigned to groups using stratified randomization. For that purpose, we used the data collected in the registration form. Table 4 shows the results of the randomization.

Table 4. Distribution of experts for experiments.

	D 1F	D1D		D2F		D2D		Total	
	M	A	M	A	M	A	M	A	
Manager	2	2	1	5	2	2	5	1	12
Analyst	7	2	4	4	2	7	4	4	17
Designer	4	3	3	4	3	3	4	4	15
Programmer	3	4	2	3	4	3	3	2	12
Scholar employer	0	0	1	0	0	0	0	1	1
Other	5	8	12	5	8	5	5	12	24
Total	21	19	23	21	19	20	21	24	97

During the experiment, we prepared a personal folder for each participant that included the use case diagrams of the task that he/she had to perform manually. In this case, participants also received an activity diagram that described one of the use cases of the diagram in detail. Besides, they were given a set of templates (4 concretely) for functional tests. Everyone got only the activity diagram that they had to execute, and they did not know what their colleagues were doing. In the first part of the experiment, each participant had to define up to four functional tests manually only with the information of the use case and the activity diagram, but using the test functional pattern included in the folder, as we previously mentioned. We concretely defined the maximal number of functional tests because in real projects, companies do not have as many resources as they would like. They have a limited one and we restricted them to four using the real experience of selected projects in the practice. The detailed description of use cases and activity diagrams is out of the scope of this paper. However, further information on the experiment can be found in our group's website.[4] After the manual exercise, our experts taught a course related to Early Testing. In that course, participants were given a detail description on the advantages of Early Testing, reduction of time and how it can help the early detection of errors in software lifecycles. During the course, participants discussed with our experts and worked with tools they would use in the second part of the experiment. After the course, our participants received a new document, the second use case and the activity diagram (different from the previous one) to be executed with NDT-Suite. Along that execution they received the functional test derived from their activity diagram only with a click. Finally, once the experiment finished, we asked them to talk about their experiences in Early Testing and its automation. This form is available in [5].

[4] http://iwt2.org/experimentacion-en-testing-temprano/

[5] This report was also presented in Opina. Obviously, obtained results were really interesting.

3.3 Running the Experiment

The right execution of the experiment was a critical aspect for our future relationship with the participating companies. For this reason, we ran a pilot in Zaragoza before the experiment. We counted with the support of the Instituto Tecnológico de Aragón[14], which invited 17 companies. During this experience, we were able to assure that our schedule was well planned, that some aspects of the Early testing course had to be changed to make them clearer for participants and that some other improvements were necessary. Additionally, three days before the experiment, we ran a new pilot in the same classroom where the real task would be conducted, with five students in the last course of the degree. These two pilots would ensure that any technical problem would be solved before the experiment. A total of 91 professionals registered for the experiment. Some of them were people who had previously worked with us in other projects (they could be consider our "costumers" in Technology Transfer projects) and some others were going to collaborate for the first time in a University project. The experiment was carried out on 4th and 5th. It was planned as a half-day activity (4.5 hours, see Table 2) developed at the School of Computer Science at the University of Seville. At the beginning of the session, we asked them to fill in a registration form that included the collection of personal data and academic and professional information similar to that in Table 1. It would guarantee that nobody made an error in his/her inscription. During the execution, we had two problems:

1. Although we had 91 people registered, only 76 participants took place in the experiment. Therefore, to mitigate this problem, the remaining people were re-assigned to groups in order to maintain the original design balanced.
2. The second problem was related to the number of computers but, it will be analyzed in detail in the next section.

As commented, information, both inscriptions and interviews at the end of the experiment was gathered through online forms, designed with OPINA; a free tool frequently used because it is very versatile and lets us explore and export results in Excel sheets. The first part of the experiment, that was to say, the manual definition of tests, was made by hand in paper. We compiled and classified them according to the starting and ending time of each participant (key for measure the first RQ). As previously mentioned, participants used NDT-Suite for the second part of the experiment consisting in the generation of tests automatically. This tool can be briefly described as a specific UML (Unified Model Lenguages) [20] profile that makes possible to apply NDT to real environments. NDT-Suite also implements all necessary transformations among models. All this framework is integrated into Enterprise Architect [4] tool. This environment, allowed each participant to be able to (i) generate tests associated with his/her use cases and (ii) produce different office documents with the structure of each test and results. That process was performed in an automatic and systematic way, but we kept the results of each participant performance to control that part of the experiment. We also recorded the time each of them spent (i.e., starting and ending time).

3.4 Threats to Validation

Our design was exposed to several validity threats such as fatigue and learning. We tried to address them by applying the following strategies:

- **Fatigue effect:** The fact that each subject has to analyze two problems means that two sessions are required to run the experiment. If these sessions are performed in close succession, subjects may experience fatigue, and in turn, they may lose effectiveness as the experimental sessions progress. To avoid this pernicious effect, the experimental sessions will take only 4.5 hours, with a coffee break to relax. Therefore, fatigue will not affect the second session, which will be developed in similar conditions to the first one.

- **Learning effect:** The source of the learning effect is the performance of the same experimental task by the same experimental subject on repeated occasions. In this experiment, each subject will study two completely different domains, and consequently the information will be unlikely to reuse from the domain-aware problem (AP1) to domain-ignorant problem (IP1). Besides, the elicitation is performed using the open interview, and subject skills are a priori unlikely to improve substantially after a mere 30-minute course (actual elicitation sessions were even shorter) and over the 2 days between sessions.

4 PRACTICAL EXPERIENCE WORKING WITH COMPANIES

This section explains how the strategy illustrated in section 3 has been embodied in a concrete experiment to get a large number of participants. It has been organized around three moments: before the running the experiment, along the experiment and after running the experiment.

4.1 Before Running the Experiment

The first idea was ***to attract our enterprise experts through their clients***. Thus, we contacted with Mr Jesús Huertas, Director General de Política Digital in Consejería de Hacienda y Administración Pública in Junta de Andalucía[2], who is responsible for the Andalusian Government IT (Information Technologies) digital policies. He is a very relevant professional, as well as a key point of contact or an important client for companies in Andalusia. The manager of our team asked for a personal meeting with him, as she knew him for several years, she presented him the experiment and asked for his endorsement. She explained in detail the importance of this kind of experience carried out among University-Public Administration-Companies, mainly in a current strategy line for IT: testing (it took her 30 minutes approximately). The endorsement that we were looking for was critical for our experiment. That meeting enables her to explain that we were also backed up by Junta de Andalucía, which had recognized our experiment as a very relevant initiative for the community. After that, as a second action ***we prepared an invitation and we designed a website with the definition of our experiment***. In that invitation, we explained some critical aspects[6]:

1. The support of Junta de Andalucía and Mr Jesús Huertas' backup.
2. A set of advantages for our participants, described below.
3. A very concrete agenda.

These aspects are themselves an attractive claim for companies for several reasons:

1. The support of Junta de Andalucía was essential because they have key clients for companies in our environment and they also set the pace and make global decisions about IT that are decisive for IT companies in Andalusia. Thus, to collaborate with the University of Seville in a Software Testing project endorsed by Mr Jesús Huertas gives companies a relevant position in the community.
2. The advantages we presented were also fundamental, as we offered two different points.
   a. First, advantages for participants. Thus, we offered a personal certificate of participation, a free course on Early Testing and a free evaluation of their performance along

[6] Once again, it is available in the website of the experiment http://iwt2.org/experimentacion-en-testing-temprano/

the experiment, where we valued their knowledge. Besides, a personal evaluation from experts in our team is also offered.

b. Second, advantages for companies. Thus, if a company participated at least with five experts, it would receive an especial certificate of participation. That certificate resulted very attractive for them, mainly for future public contracts in Andalusia.

3. Defining a very concrete agenda was also a relevant aspect. People in companies are frequently very busy; to spend time in an experiment is often very difficult for them, so an agenda may help them to close and reserve a specific period of time for the experiment. Besides, we set three different sessions, with the idea that they could adapt their participation in the project to their professional agendas.

However, our experience with companies lets us know that an official invitation is not enough, as everyday they receive a lot of emails and some of them, despite their attractive designs, are usually classified as spams. Consequently, if we only sent the invitation, it would probably be discarded and considered as a spam. Thus, with the official invitation and as a third strategy, ***we sent personal emails to companies' staff***. The fact of sending the invitation helped us to learn that the two points below were very important:

1. To send individual emails with a personalized text for each recipient. We did not send our mails to companies in general. They were sent to concrete people (our own contacts, our contacts' contacts or even recommended addressees by Junta de Andalucía). Each email was different and each of them attached the personal invitation. where we asked participants to resend it to other colleagues.
2. To contact people through phone calls, WhatsApps, social nets or other similar platforms of communication, as we considered that, despite this personal mailing, people quick rhythm of life would make them discard the email without reading it..

Obviously, those kinds of calls were very expensive in resources. In our experiment, it was the director of the team who mainly executed that task, which was also a very relevant aspect; for companies, if it is the boss who makes the phone call, the fact will acquire more relevance.

Table 5 below numbers the mails, WhatsApps and phone calls made for inviting people.

Table 5. The number of contacts made by the IWT2 leader during the call is presented [7] .

Number of sent mails	337
Number of direct phone calls related with the experiment	131
Number of sent WhatsApps directly related with the experiment	350[8]

However, we considered that this personalized dedication was not enough. It was key for our experiment to attract different profiles from SMEs (Small and medium enterprise), big companies and public companies. Besides, we claimed for different levels of experts: senior and junior people. Thus, we also asked for a face-to-face interview in case of key participants. Consequently, as a fourth strategy, we demanded face-to-face meetings with some companies' members of the staff who were very important for our experiment. Thus, for instance, with ATOS[1], whose international testing factory is based in Seville, or with IECISA[13][9], we held a personal meeting with the direction of the factory to invite them to participate in our project.

[7] This table only presents specific calls and mails for this experiment. In other meetings or environments, the experiment was also referred but it was not presented in this paper.

[8] It is a reference since a conversation can include several messages. It refers to the number of conversations.

[9] IECISA: Informática de El Corte Inglés

Table 6 offers the number of meetings that the leader of our team held with different companies, grouped into SMEs, big companies or public companies. We deemed to have such meetings in the companies' headquarters as something relevant, despite the cost of displacement,, as some of them took place even out of Seville.

Table 6. The number of face-to-face meetings held by the IWT2 leader during the presentation of the call.

SME visits	Public companies visits	Big companies visits	Total
4	4	5	13

We can assure that all visited companies participated in the experiment. Some of them even asked our team to replicate the experiment in the own company. Thus, as we will conclude in the last section, one-to-one meetings were very expensive, but a very good and profitable strategy.

Table 7. Total number of participant.

	Senior	Medium	Junior	Testing Experts	Gender		**Total**
					men	women	
SME	2	12	2	1	15	1	**16**
Big Companies	26	15	14	34	45	10	**55**
Public companies	8	5	3	0	12	4	**16**

Finally, to finish with our experiment arrangements, we defined an online registration form described in Table 1. ***The strategy of having an online registration form was also a suitable idea.*** That form would provide us with relevant aspects as follows:

1. We would follow the number of participants and the influence of our calls. Thus, if a company or someone was not inscribed, we would insist with more calls or mails.
2. We included some demographical and personal data (like experience or level of test knowledge, for instance) that would help us to define the scope of our experiment, as it will be further explained in the next subsection.
3. Each participant would select the best section for him/her. That was critical to assure that we would assume the number of participants in each section.

After carrying out the task of contacting people, Table 7 shows the final number of participants.

4.2 Along the experiment

Experiments were executed on November 4th and 5th in three different sessions; one on Wednesday afternoon and the other two on Thursday morning and afternoon, respetively. We did not define the proceeding of those sessions, as in fact they were defined in collaboration with companies, particularly with bosses along the face-to-face meetings previously held with them. ***Coordination with companies*** was a successful idea in order to guarantee the participation of our experts. Another important issue was the preparation of the class. As we commented, we executed a similar project in Zaragoza some weeks before. Besides, on Monday before the agreed date for the experiment, seven people executed it in the same classroom where it would take place, just to check if the software was right, if computers were available, if web connection was running properly, and some other

concerning aspects. ***This test was considered as a very interesting strategy since it gave us time to repair some computers with software problems*** (just technical problems when installing Enterprise Architect or with the internet connection). It is also remarkable that some days before the experiment, we ***sent personal emails to each participant as a reminder of the timetable, the agenda and the some other arrangements related to his/her participation in the project***. Those emails aimed not only to remind, but also to confirm their attendance to the meeting.

As we introduced, we worked with two different systems and with two kinds of use cases for each of them: four examples in total. Then, we divided cases with on-line supporting among our participants, in order to assure their good distribution. The email enabled us to guarantee that we had a right number of examples and a right distribution. However, we also considered other elements, not only based on the methodology for software experiments, but also related to the corporative and public image of our group and our experiment. They were mainly these three listed as follows:

1. Each participant received a folder with personalized documentation for the experiment. In fact, we were sponsored by Fidetia[9], which provided folders, pens and notebooks, by offering a very serious and interesting view of our experiment.
2. The coffee break was seen as a relevant moment that allowed improving networks and relations among participants. We planned to offer a coffee, although we then considered that as the experiment involved a short period of time, it was not a good idea to move our participants from one place to another. Therefore, we used a classroom with a special area for the coffee break.
3. Our participants valued time as a very positive aspect and basically, as one of the most relevant learned lessons in our experience: to **be on time was essential to collaborate with companies,** because time means money and they usually have very close and complex agendas.

Nevertheless, our experiment not only provided good results; during the execution we addressed two important negative aspects. The first one was related to the number of PCs. We reserved a classroom in the School of Computer Languages and Systems and we had to face up a very simple problem. In the form designed to book classrooms, the School recommended the maximum capacity of each classroom.. Thus, we thought that this figure involved the total number of computers in the classroom as well. In consequence, the place was designed with one computer for two people and we had to divide participants in two different groups. It was important in the session of Wednesday morning since it brought together the largest number of participants. It was really a misfortune that could be avoided, as we performed a test on the previous Monday and we did not realize it. Moreover, we had to change the execution at that moment, and we even used our personal laptops to decrease the effect that issue would cause on the timetable of the experiment. As a result, we were delayed 30 minutes approximately.

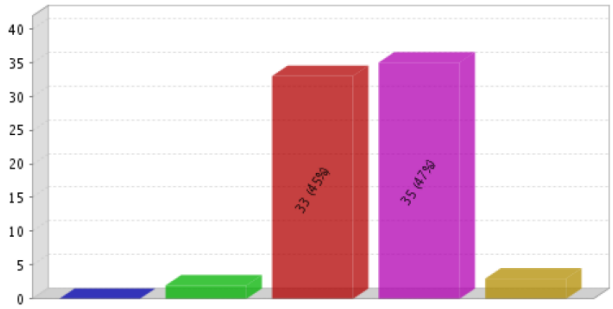
Another drawback was a conceptual decision that we made regarding the definition of the experiment. We considered that it could be interesting to conduct the experiment with people of different level of knowledge: juniors, seniors, managers, and so on. The results obtained with that decision were really good but they entailed a considerable problem in the execution of the experiment, because along the Early Testing course we had a very heterogeneous audience. We reduced the effect of this problem by involving in the project many team members (more than 15 people participated in each session), who assisted and helped some of the youngest participants[10].

[10] It is very important to stick out that help did not mean conducting the experiment, but only solving problems or clarifying concepts related with metamodels, transformations or even

4.3 After Running the Experiment

After conducting the experiment, we also followed some strategies to demonstrate our participants the positive effect that taking part in university experiments may entail.
Obviously, by gathering information, we learned a lot of aspects that we had never thought before about the use of Early Testing as well as how valuable experts considered our tools. However, the analysis of these conclusions is not the aim of this paper. As we introduced, we intend to analyze which strategies are suitable to improve communication with companies in order to get their involvement in software experiments.

Even though we are not going to present the conclusions of our research, Figure 1 presents an interesting graph obtained from OPINA with the answers to a very specific question asked to each participant in the last questionnaire of the experiment: in your view, is this kind of experience developed between university and companies interesting?



Blue: It was not interesting	0%
Green: I consider I did not learn anything	2%
Red: Yes, but I wanted to get more learning	45%
Pink: Yes, I liked it and I learned very much	47%
Yellow: It was wonderful	4%

Fig. 1. Satisfaction percentage

As it can be concluded from this chart, a large number of our participants considered the experiment as very interesting as they thought that collaboration between entities could be very suitable for improving their jobs. They also stated that it was an profitable working day that offered very relevant solutions to enhance their own businesses. For that reason, and taking into account that 98% of participants judged the experience as very relevant for their jobs, we deem it necessary that the research community improves this kind of collaborations.

After the experiment, we also continued applying strategies to keep the interest of our partner companies and participants. Firstly, after the execution of the experiment, we sent **a personal email with the global figures taken out from the experiment** (number of participants, companies, and results, for example), just to demonstrate the significance of the results. **We also sent a letter to Junta de Andalucía and we published a especial news about the success of the experiment in our university social networks**.

Two weeks after conducting the project, each participant received his/her **certificate of participation via email** where they were offered the possibility to be sent the original one. Then, the

functional test definitions. Our team did not influence the experts' execution of the experiment.

team was assessing the results throughout the next three months (December, January and February). Currently, this evaluation has just finished and next week we will send our participants a new email, giving them the possibility to know their results. We guaranteed the confidentiality of the results in the experiment. In consequence, we cannot give information about each participant and his/her performance, but we can reveal their personal results and positions in comparison to other participants. It means that we are going to meet each participant who deems it necessary and, depending on his/her profile, we will compare his/her results with other participant's results with the same profile as well as will give them some advise under our consideration. Besides, they can **freely download examples, tools and manuals** used along the experiments and they can have a direct connection with the research team in case that they want to know more about Early Testing. We have established a direct link to companies as well. This experience has also strengthened the communication with companies we had worked before, by means of demonstrating that the experiment is also relevant for them. For this reason, the results have led to a new phase. We are closing face-to-face meetings with manager teams at companies by offering them a detailed evaluation of the current situation of their experts, and we will freely provide them with a theoretical evaluation of their present situation. Obviously, we have to be very careful with these meetings, as we do not pretend to be conceited. Companies know very well their own businesses, thus, we will only offer a constructive opinion according to our experiment's results.

5 FINAL SUMMARY OF SUCCESSFUL STRATEGIES

To finish and summarize, Table 8 presents learned positive lessons for each phase of the experiment (those failed are presented in the test).

Table 8. Summary schema.

preparation	• To get the support of companies' costumers is essential to attract their interest for the experiment. • To define a very concrete call, with clients sponsoring, as well as a very concrete agenda is critical. The call has to include clearly advantages for companies. • To offer personal invitations (by means of personal emails, phone calls, WhatsApps and some other means of communication). • To hold face-to-face meetings with manager teams and to engage them in the project. • To keep time and schedule arrangements as carefully as possible. Time means money for companies, thus time investments have to be well described.
execution	• To agree with companies on the days and hours for the execution, even if several sessions or "extra hours" need to be scheduled, for instance, on Friday evening or at weekends. • To execute as many tests as possible to check that there is not any technical problem or problems with the infrastructure or facilities. • To prepare personal documentation for each participant. Each participant is important for the experiment, and this is what they should be demonstrated. • To be very strict with time, as it is one of the most critical aspects. • To consider time for coffee as a key point to promote personal relations and future collaborations networks. Companies are also interested in improving and increasing their networks.

analysis	• To send information to participants. The experiment has to offer results and our participants has to know it. • To be sure that you offer all the advantages that you promise in your experiment. • To try to have face to face meeting with companies to learn even more. • To try to explote the network that you can get with an experiment like this us. It is a good way for future works and transference projects.

6 CONCLUSIONS AND FUTURE WORK

This paper has presented the global view of a software experiment executed at the University of Seville in collaboration with the Politecnical University of Madrid and Instituto Tecnológico de Aragón. The paper does not focus on presenting the experiment. In fact, the project has provided interesting feedback to go on researching. However, one of the most relevant aspects of this study is amount of people from companies engaged to collaborate in the experiment. There were 97 experts registered and finally only 76 participated, with a total of 32 companies involved. As the literature demonstrates, to get this number of participants is very difficult. Therefore, this paper tries to present which strategies were used to get this goal. As a future work, we would like to publish and to get information about the experiment and obviously, we aim to improve our tools. Besides, after the success of this first experience, we are looking forward repeating it with other areas of NDT like quality, project management or requirements management. Besides, many companies that took part in the experiment asked to repeat it with all their employees at their own offices or facilities, what makes us be very confident. For instance, companies from Zaragoza, which participated in the prototype, demanded to repeat the experience at their own headquarters.

Finally, this paper will conclude with a global consideration. This study has confirmed that communication and collaboration with companies is possible. Nevertheless, communication has to be in two directions. If we aim to involve companies in our experiments, we will have to consider their situation and availability and we will simplify the process as much as possible, without being detrimental to our scientific method.

ACKNOWLEDGEMENTS

This research has been partially supported by the Megus project (TIN2013-46928-C3-3-R) and by the SoftPLM Network (TIN2015-71938-REDT) of the Spanish Ministry of Economy and Competitiveness.

REFERENCES

[1] ATOS Origin. http://atos.net/en-us/home.html. Last accessed 03/2016.
[2] Consejería de Hacienda y administración Pública. Junta de Andalucía. http://www.juntadeandalucia.es/haciendayadministracionpublica/. Last accessed 03/2016.
[3] Dieste, O., Juristo, N., & Martínez, M. D. 2013. Software industry experiments: A systematic literature review. *In Proceedings of the 1st International Workshop on Conducting Empirical Studies in Industry (pp. 2-8).* IEEE Press.
[4] Enterprise Architect. SparxSystems. Enterprise Architect. Website: www.sparxsystems.com.au. Last accessed 03/2016.

[5] Escalona, M.J. Models and Techniques for the Specification and Analysis of Navigation in Software Systems. 2004. *Ph-thesis. University of Seville*, Spain.
[6] Escalona, M. J., Gutiérrez, J. J., Villadiego, D., León, A., & Torres, J. (2007). Practical experiences in web engineering. In *Advances in Information Systems Development* (pp. 421-433). Springer US.
[7] Escalona, M. J. & Aragón, G. 2008. NDT. A model-driven approach for web requirements. *Software Engineering, IEEE Transactions on*, *34*(3), 377-390.
[8] Escalona, M. J., García-García, J. A., Mas, F., Oliva, M., & Del Valle, C. (2013). Applying Model-Driven Paradigm: CALIPSOneo Experience. In *CAiSE IT* (pp. 25-32).
[9] Fidetia. www.fidetia.es. Last accessed 03/2016.
[10] García-García, J. A., Ortega, M. A., García-Borgoñón, L., & Escalona, M. J. 2012. NDT-Suite: a model-based suite for the application of NDT. In Web Engineering (pp. 469-472).
[11] Gutiérrez, J. J., Escalona, M. J., Mejías, M., Torres, J., & Centeno, A. H. (2008, June). A case study for generating test cases from use cases. In *Research Challenges in Information Science, 2008. RCIS 2008. Second International Conference on* (pp. 209-214). IEEE.
[12] Hotel Ambassador. iwt2.org/actividad-grupo/investigacion/resultados/ndt/aprendiendo-ndt/. Last accessed 03/2016.
[13] IECISA. www.iecisa.com/. Last accessed 03/2016.
[14] Instituto Tecnológico de Aragón. http://www.itainnova.es/. Last accessed 03/2016.
[15] Jedlitschka, A., Juristo, N., & Rombach, D. 2014. Reporting experiments to satisfy professionals' information needs. *In Empirical Software Engineering*, 19(6), 1921-1955.
[16] Jedlitschka, A., & Pfahl, D. 2005, Reporting guidelines for controlled experiments in software engineering. *In International Symposium on Empirical Software Engineering*, (pp. 10-pp). IEEE.
[17] Juristo, N., Moreno, A.M. 2001. Basics of Software Engineering Experimentation .
[18] Martínez-García, A., García-García, J. A., Escalona, M. J., & Parra-Calderón, C. L. (2015). Working with the HL7 metamodel in a Model Driven Engineering context. *Journal of biomedical informatics*, *57*, 415-424.
[19] Opina. opinahq.com/. Last accessed 03/2016.
[20] Object Management Group. Unified Model Language. V2.5. www.omg.org/spec/UML/2.5. Last accessed 03/2016.
[21] Punto de informacion del Flamenco. www.juntadeandalucia.es/cultura/ centroandaluzflamenco/Pif/index.php. Last accessed 03/2016.
[22] Robson, C., & McCartan, K. 2016. Real world research. *Wiley*.
[23] Runeson, P., & Höst, M. 2009. Guidelines for conducting and reporting case study research in software engineering. *In Empirical software engineering*, 14(2), 131-164
[24] Sjøberg, D. I., Hannay, J. E., Hansen, O., Kampenes, V. B., Karahasanovic, A., Liborg, N. K., & Rekdal, A. C. 2005. A survey of controlled experiments in software engineering. *In IEEE Transactions on Software Engineering*, 31(9), 733-753.
[25] Topçu, O., Durak, U., Oğuztüzün, H., & Yilmaz, L. (2016). Model Driven Engineering. In *Distributed Simulation* (pp. 23-38). Springer International Publishing.
[26] Vegas, S., Dieste, Ó., & Juristo, N. 2015. Difficulties in running experiments in the software industry: experiences from the trenches. *In Proceedings of the Third International Workshop on Conducting Empirical Studies in Industry* (pp. 3-9). IEEE Press.
[27] Web Engineering and Early Testing Group. University of Seville 2016. www.iwt2.org. Last accessed 03/2016.
[28] Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., & Wesslén, A. 2012. Experimentation in software engineering. *In Springer Science & Business Media*.
[29] Spector, A. Z. 1989. Achieving application requirements. In *Distributed Systems*, S. Mullender, Ed. ACM Press Frontier Series. ACM, New York, NY, 19-33. DOI= http://doi.acm.org/10.1145/90417.90738. Last accessed 03/2016.