

Difficulties in Running Experiments in the Software Industry: Experiences from the Trenches

Sira Vegas
Universidad Politécnica de Madrid

Background

- Laboratory experiments are common practice in SE
- Laboratory experiment = Simplified reality
 - Students vs. professionals
 - Toy software vs. real systems
 - Exercises vs. real projects
 - Just learned vs. knowledge & experience
- Laboratory findings **MUST** be generalized through other types of experiments: e.g. experimentation in industry

Experimentation in the Sw. Industry: State of the Practice

- Most controlled SE experiments are run in academia
- Conduct experiments in the software industry is challenging: few experiences
- Previous attempts at running experiments in the software industry:
 - NASA SEL-University of Maryland
 - Daimler – Ulm University
 - Simula

Our Approach

	# Companies	University	Replication
SEL-UMD	Single	Single	Not systematic
Daimler-Ulm	Single	No	No
Simula	Multiple	No	No

- Run the same experiment in several companies and several universities

	# Companies	University	Replication
Our approach	Multiple	Multiple	Systematic

Experiment Description

- RQ: How does TDD compare to ITL regarding: amount of work done, code quality and developers' productivity?
- Treatments: TDD vs. ITL
- Response variables
 - Amount of work done: Tackled user stories
 - Quality: Quality of tackled user stories
 - Productivity: Amount of work successfully delivered
- Tasks:
 - MarsRover
 - Modified version of Robert Martin's Bowling Score Keeper
 - MusicPhone
- Experiment run in either Java or C++

Concept Warmly Welcomed

- Company decisions are usually based on:
 - Marketing speak
 - Recommendations of a consultant
- The idea of having a means to objectively and quantitatively evaluate technologies and methods was appealing

But...

Identified Difficulties: Company Involvement

- **D1. Concept tough to grasp**
They do not see how the idea will materialize
- **D2. We need more than one subject**
Confusion with single-subject study
- **D3. Experiment as a free training course**
Win-win strategy. Both parties get a benefit

Course-experiment bound: a bad marriage for us

- Subject are not proficient on the task
- Causes trouble with participants:
 - Must accept some differences from a regular course
 - Reluctance to training
 - Non-constructive discussion
 - Pressure on trainer
- Subjects' perception on training has an effect on motivation

Identified Difficulties: Experiment Planning

- **D4. Choose experiment topic**

Most companies hardly seemed to care which topic was investigated

- **D5. Choosing experimental tasks**

Companies did not provide us with experimental tasks

- **D6. Getting experimental subjects**

Innovation manager does not have the power to enrol people in a course. Internal organization critical

- **D7. Selecting a design: few degrees of freedom**

Constrained by small number of participants (within-subjects), and course as experiment (AB design)

Identified Difficulties: Experiment Execution

- **D8. Facilities might not be available**

Harder to gain access to computers

- **D9. Privacy and security issues**

- Impossibility to install specific instrumentation on computers => virtual machines
- Access to resources denied: network, printing/storing data, access to rooms only at given times

- **D10. Company technology unsuitable**

All material in Java and Junit. Extra work porting tasks, test cases, etc.

- **D11. Dropouts**

Due to proximity between working and experimental environments, subjects skip parts of the course

Identified Difficulties: Data Analysis and Reporting

- **D12. Missing data**

Due to dropouts. Critical for within-subjects experiments

- **D13. Large variability in data**

Larger than in students. Could be due to either differences in background or motivation. They do not perform better than students. Only high-performing ones

- **D14. Rush for results**

As a result, we made mistakes during data measurement, and analyses had to be repeated several times. Took us longer than expected

- **D15. Reporting must be adapted**

Managers do not necessarily have knowledge of statistics/experimental design. Simple and visual representations

Conclusions

- Difficult to materialize a very welcomed concept
- Industrial environment imposed constraints
- Professionals were troublesome, under motivated, and did not perform better than students
- Results reliability could be influenced by specific characteristics of data: missing, variability, etc.
- Reporting used in journals not appropriate