## Experimental Software Engineering (ESE)

Group: <a href="http://eseg.uniroma2.it/">http://eseg.uniroma2.it/</a>

### ESE

## 6E1. Example of Experiment Definition

## **Practical Example** Comparing Code reading and Functional testing with focus on Effectiveness and Defect **Detection Rate** for OO event-driven Java software with sophomore students for research.

# Experiment Definition: Dictionary

- Code Reading (CR): is intended to detect software defects by entering directly to software artifacts. Once a defect is located, the location is reported; usually, the defect is also explained.
- Functional testing can (only) detect failures (unexpected behaviors owing to the existence of faults).
- Functional Testing & [Defect] Identification (FTI): once a failure is detected, faults are identified in the code, e.g. they are located and motivated.



DICII – UniRoma2 – Giovanni Cantone

#### Experiment Definition: Dictionary . CR vs. FTI

### What Testers Have To Do

- Code Readers are given the program specification (intended function), some design documents (e.g. class diagram), and source code (derived function):
  - In order to detect defects, they read the code, understand behaviors, and compare the intended functionality with actual derived one. Eventually they define the class and line of code that could cause failure.
- Functional Testers are given the program specification and the executable version of source code.
  - They start detecting failures by comparing the intended functionality with the software actual behavior.
  - In order to identify defects, they pass to read code and possibly some other design artifacts (e.g. class diagram). Eventually they define the class and line of code that cause that failure.

#### **Experiment Definition**

### **Experiment Conjecture**

- Baseline: There are defects of kinds DA, DB and DC, which remains in the code of software games but should not.
- Test manager conjecture: Code Reading helps significantly more (e.g., +5% in the average) than Functional testing in detecting defects of those kinds.
- Test manager motivation: Reading was observed to perform better than Functional testing for some types of defects with expert testers (e.g. see UMD-CS experiments as reported by TSE).



Q1: Should we introduce code-reading in the company process for the development of software games?

#### Experiment Definition ESEngineer Observation

Due to the specific type (Graphic software) of the investigated objects, FTI testers, similarly to users, interact with objects through user interfaces, rather than other means.

#### Experiment Definition

### **ES Engineer Reasoning**

#### Because :

The visibility of failures should be very high, and

Time strictly necessary to detect such a failures should be very limited,

As a Result :

The FTI failure detection phase should have very minor effects.

# Experiment Definition ESE Answer

In theory, we should not introduce CR in our Software Processes. However, we need to give empirical evidence to such an answer.

#### Explanation

FTI reading phase should perform better than CR with respect to the common task of identifying software defects in the implementation of games and any kind graphic-interfacebound software; that is because it is easy detect failures in the behavior of that kind of software, the FTI reading is driven by known failures and software specifications, while the CR is driven by the software specifications only. This should make a major difference.

#### **Experiment Definition**

## **The Experiment Idea**

To match two fault detection techniques through an experiment in order to evaluate their performance with respect to different variables such as: ad-hoc for code reading techniques, functional testing & fault identification for testing techniques, time consuming, and several types of seeded faults. **Informal Hypothesis** (ES Engineer): The current technique, FTI, should perform better or not worse than CR for any type of defects for graphic-bound software, whatever the experience level of test people might be.



### Experiment Definition Response Variables

**Effectiveness:** Number of faults detected, in the average (Real number, Ratio scale). **Efficiency**: Number of faults detected per unit of detection time, in the average (Real number, Ratio scale).

**Experiment Definition Direct Measurements** Known Faults: seeded faults plus new detected faults; (integer, Absolute scale). **Per subject:** Number of Positive Faults **Detected:** (integer, Absolute scale). **Fault Detection Time**: (sec's, Ratio)

### Experiment Definition Indirect Measurements for the Response Variables

**Technique Effectiveness** 

**Technique Efficiency** 

totalPositiveFaultsDetected

totalKnownFaults

where:

totalKnownFaults = (totalSeededFaults + newFound)\* NumberOfSubjects

totalPositiveFaultsDetected

totalTestingTime

# Experiment Definition GQM

To analyze Code reading and Functional testing & fault-Identification, for the purpose of comparing their performance with respect to effectiveness and efficiency of testing event-driven OO Java software games for defect detection and classification, in the context of a technology-bound software academic lab and students of three academic levels of experience and knowledge, from the point of view of an ESE applied-research group, in the aims of transferring the resulting experiment process and guidelines to industrial settings in the mid-term.

#### Experiment Definition Experiment context characterization: Subject & Objects

•		# Objects	
		One	More than one
	One		
#			
Subjects			
per	More		Blocked
Object	than one		subject-object study