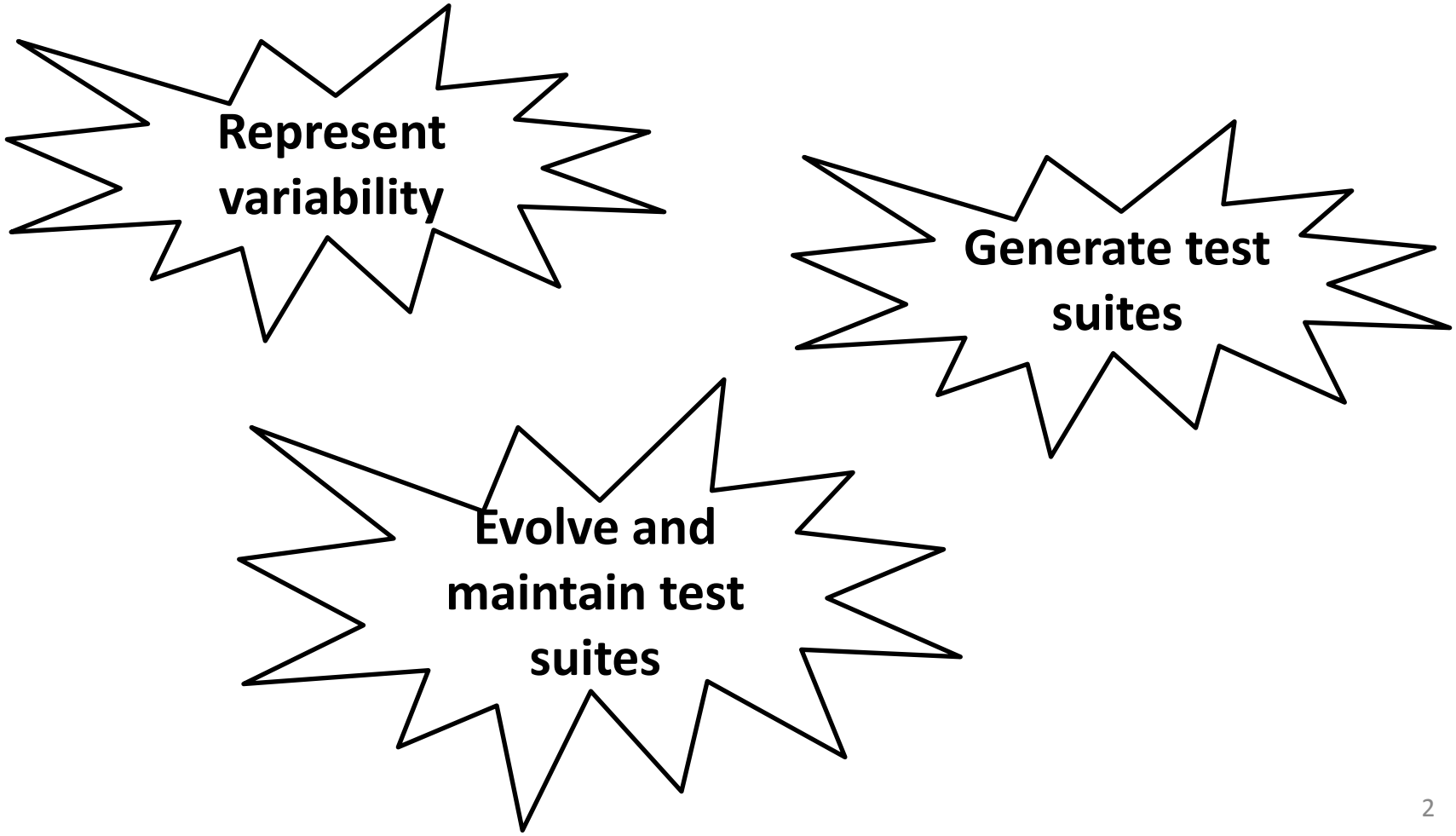
Comparing Different Test Strategies for Software Product Lines

Paola Accioly (prga@cin.ufpe.br)

Advisor: Paulo Borba (phmb@cin.ufpe.br)
Co-advisor: Rodrigo Bonifácio (rba2@cin.ufpe.br)

How to specify black box test cases for software product lines?

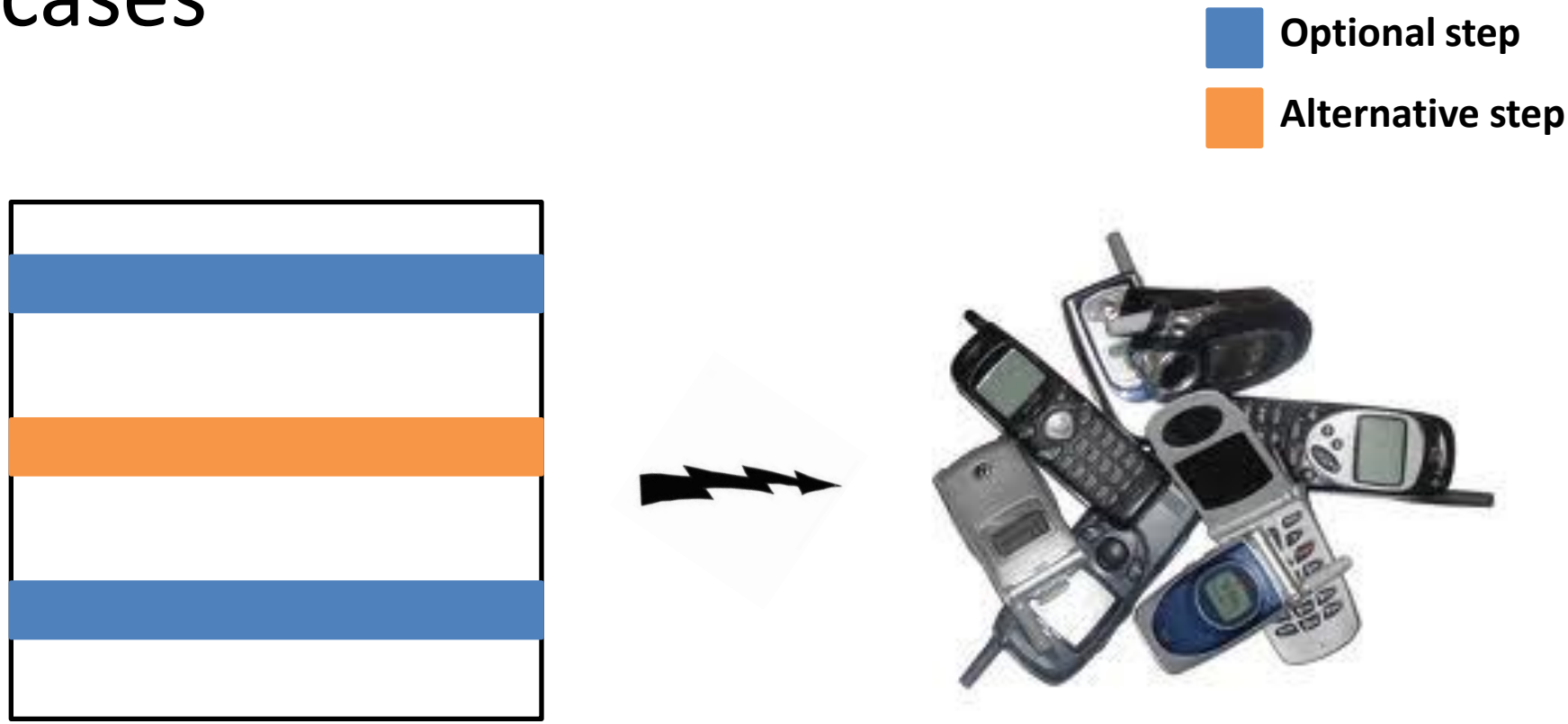


Recently some techniques have been proposed…

However, they still require further evaluation.

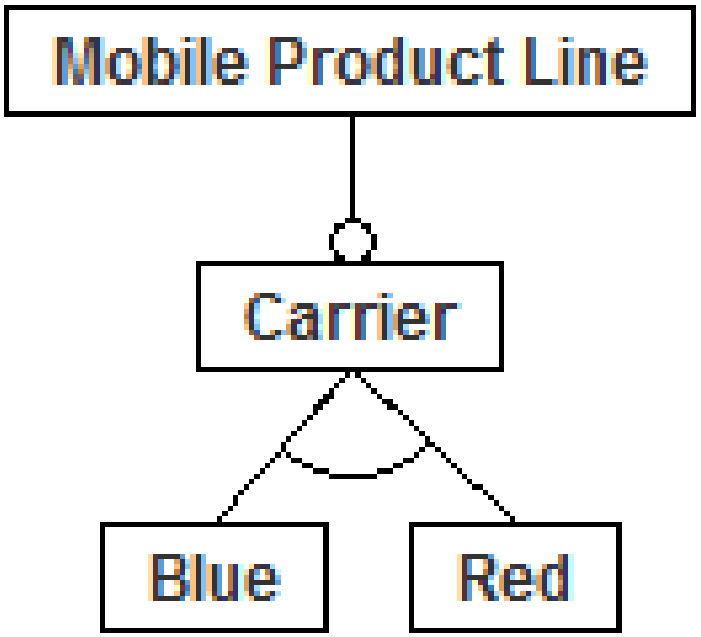
Consequently, the industry is not encouraged to invest in adopting such techniques.

One possible solution, that we have observed in a real test execution environment, is the use of generic test cases



However, using generic test cases can bring some problems to the test execution process. Let’s take a closer look at one example!

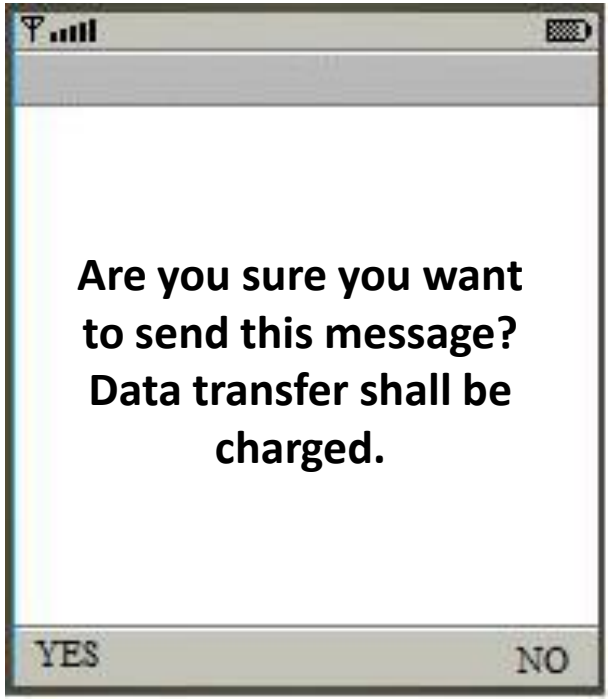
Mobile product line



Test case: user sends MMS with picture attached

Step Nº	User Action	System Response
1	Go to Main Menu	Main Menu appears
2	Go to Messages Menu	Message Menu appears
3	Select ‘Create new Message’	Message Editor screen is shown
4	Add Recipient	Recipient is added
5	Select ‘Insert Picture’	Insert Picture Menu is shown
6	Select Picture	Picture is Selected
7	Select ‘Send Message’	Message is correctly sent

Test Case



Product Behavior

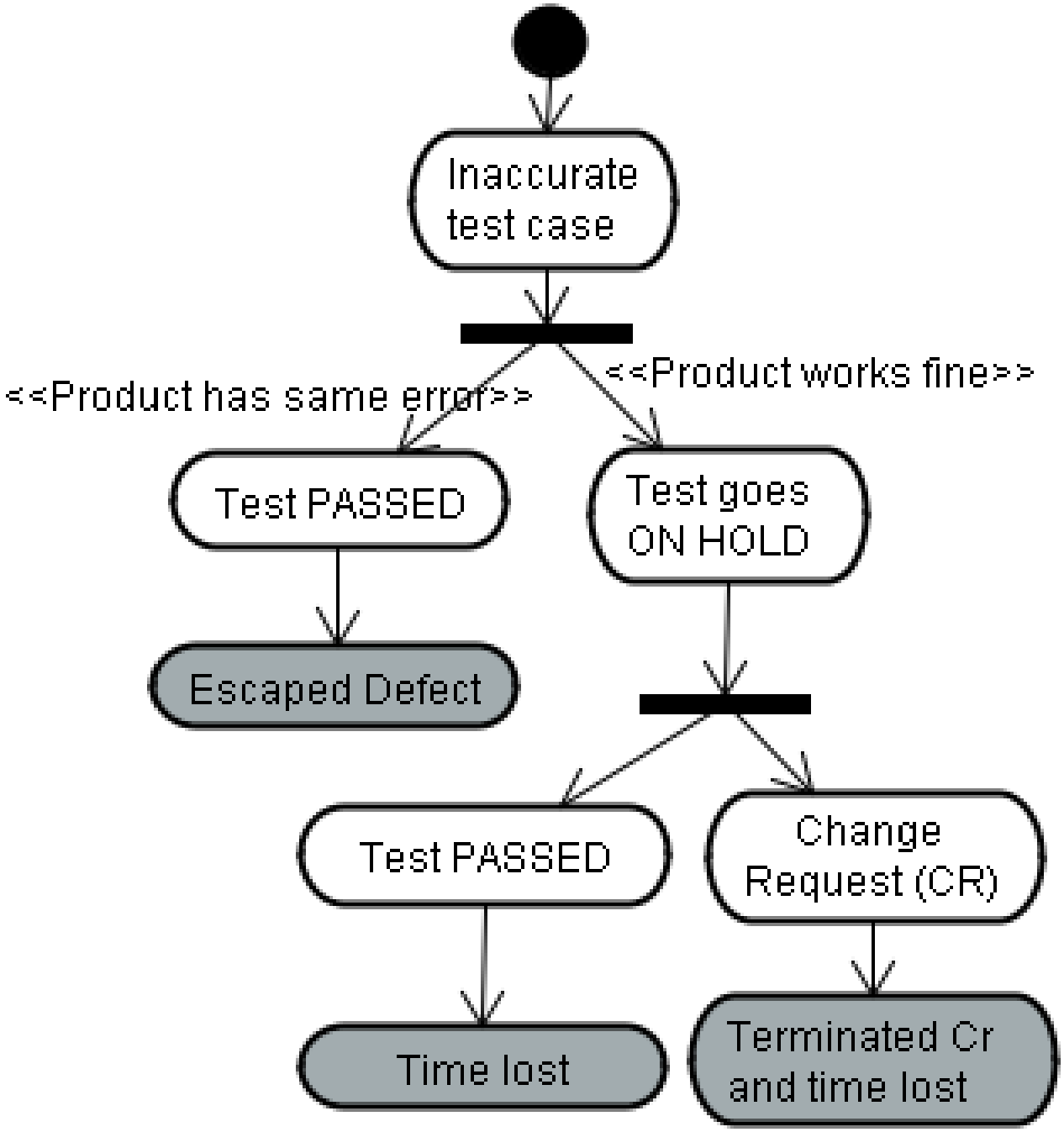
Specific test case for products configured with the Blue carrier feature

Step №	User Action	System Response
1	Go to Main Menu	Main Menu appears
2	Go to Messages Menu	Message Menu appears
3	Select ‘Create new Message’	Message Editor screen is shown
4	Add Recipient	Recipient is added
5	Select ‘Insert Picture’	Insert Picture Menu is shown
6	Select Picture	Picture is Selected
7	Select ‘Send Message’	**Dialog is shown: ‘Are you sure you want to send this message? Data transfer shall be charged’. Options are: ‘Yes’ or ‘No’**
8	**Hit ‘Yes’**	**Message is correctly sent**

To sum up, generic test cases may present…

- ✘ Fewer steps than necessary
- ✘ Wrong parameters values like icons and labels
- ✘ More steps than necessary

And what are the consequences of these inaccuracies?


Inaccurate test case
<<Product has same error>>
<<Product works fine>>
Test PASSED
Test goes ON HOLD
Escaped Defect
Test PASSED
Change Request (CR)
Time lost
Terminated Cr and time lost

Problems

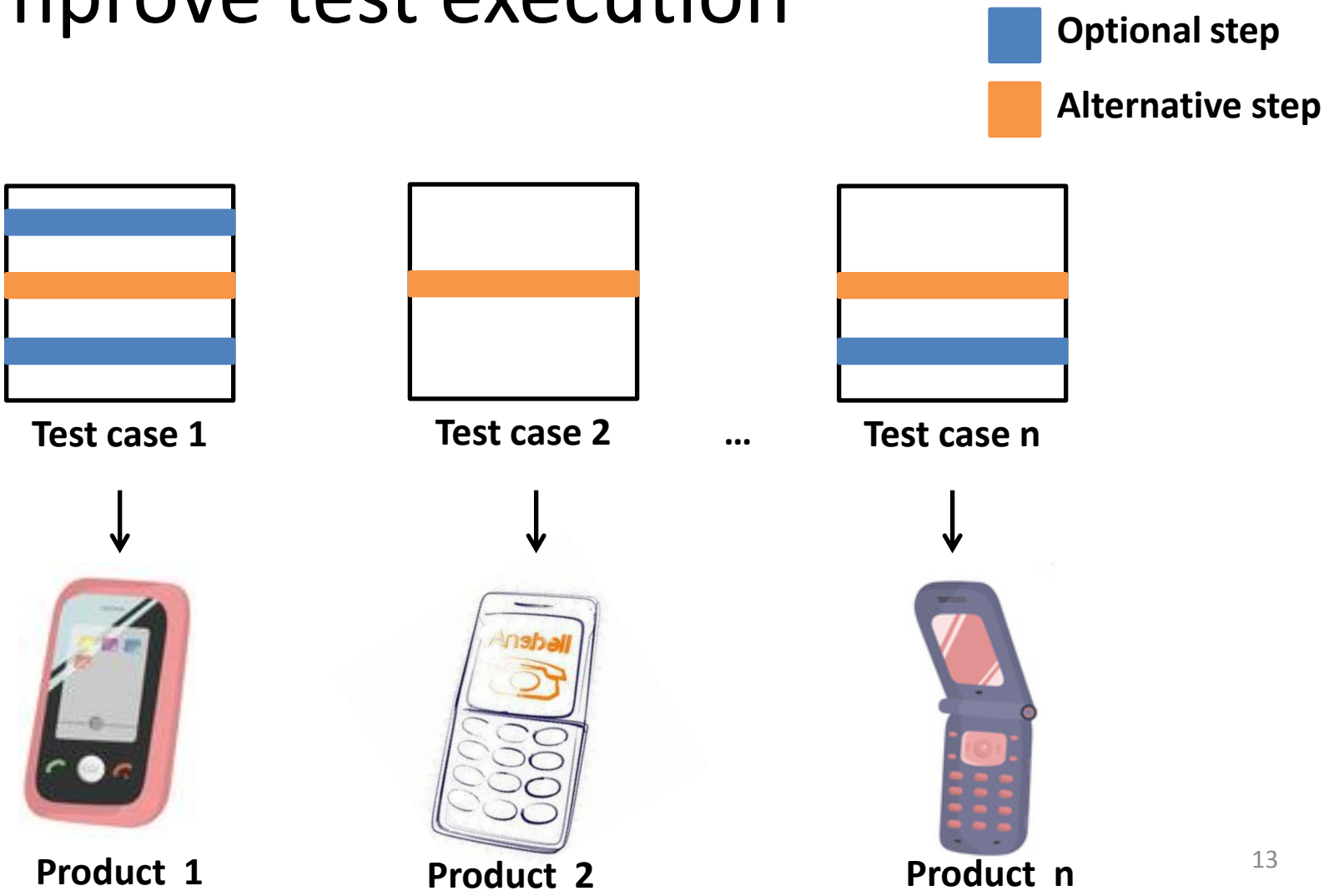
Escaped Defects affect directly the products quality

Time lost and a high rate of terminated CRs affect test-cycle productivity

Error prone

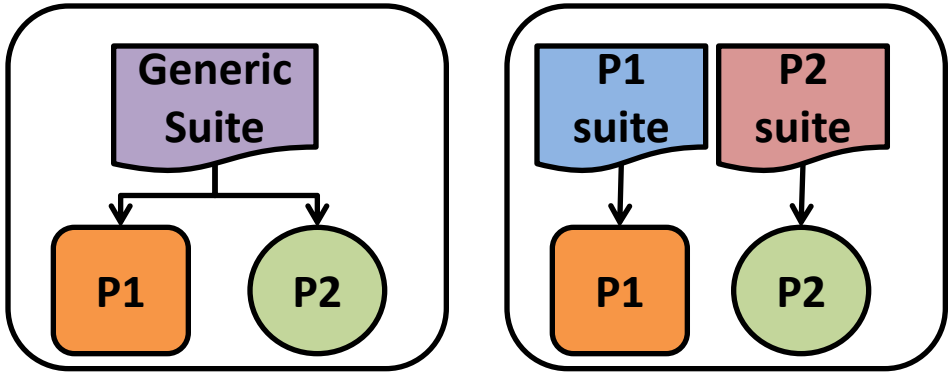
Low productivity

Having specific test cases obtained by test derivation techniques might help to improve test execution

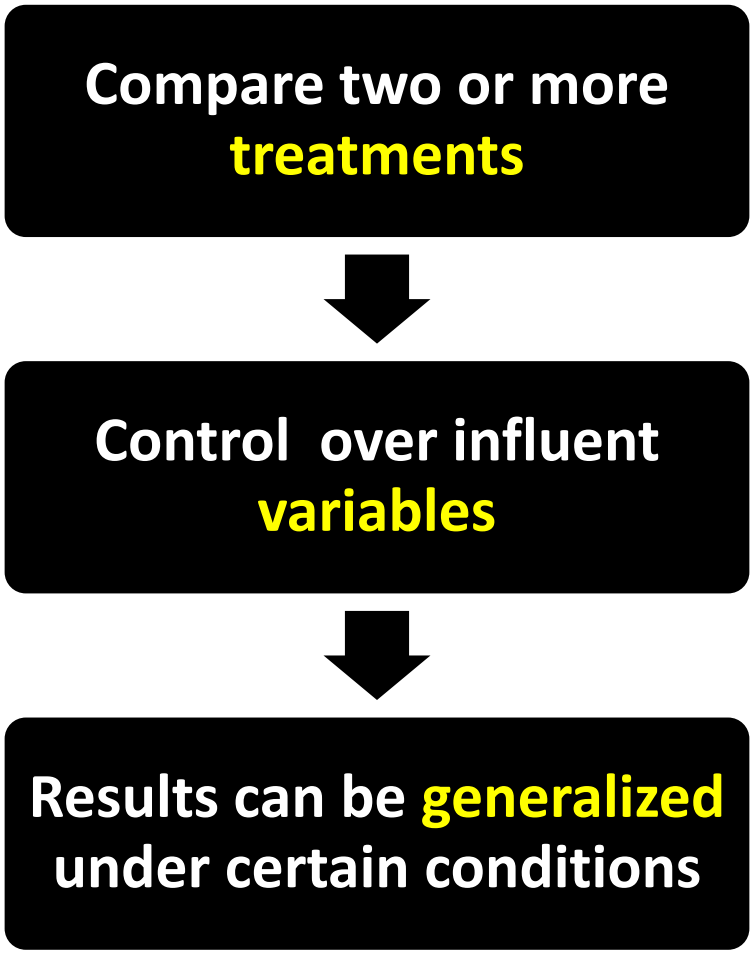


Our proposal

To compare both techniques (generic vs. specific) to investigate their impact from the point of view of the test execution process



Controlled experiments



GQM

Goal:

Evaluate two different test case design techniques for SPL (GT vs. ST) to analyze their impact from the point of view of the test execution process

Questions:

Does the ST reduce the test execution effort compared to test execution effort using the GT ?

Does the ST reduce the number of terminated CRs rates compared with the GT?

Metrics:

Test execution time

Number of Terminated CRs

The Latin square design

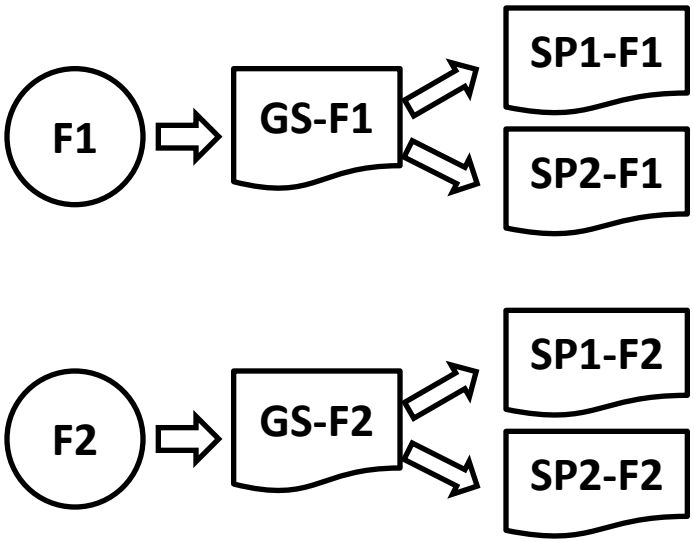
	Feature1	Feature2
Subject1	**GT**	**ST**
Subject2	**ST**	**GT**

	Feature1	Feature2
Subject3	**GT**	**ST**
Subject4	**ST**	**GT**

...

GT– Generic Technique
ST– Specific Technique

Test suites design



Differences Between Test Cases

Generic Test

User Action	System Response
Verify the options for report generation format	The options (pdf, bibtex) are available.

Specific Test

User Action	System Response
Verify the options for report generation format	The option bibtex is available.

Experiment Operation

DAY 1	DAY 2	DAY 3
Training and dry-run	Latin square first column (Feature 1)	Latin square second column (Feature 2)

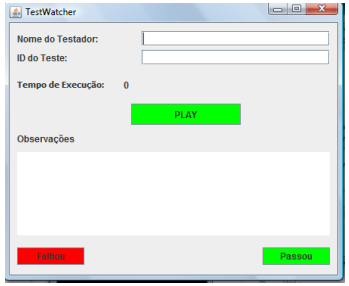
Five experiments were executed but in this presentation we'll focus on the fourth and fifth execution

FOURTH EXPERIMENT

Tools and participants

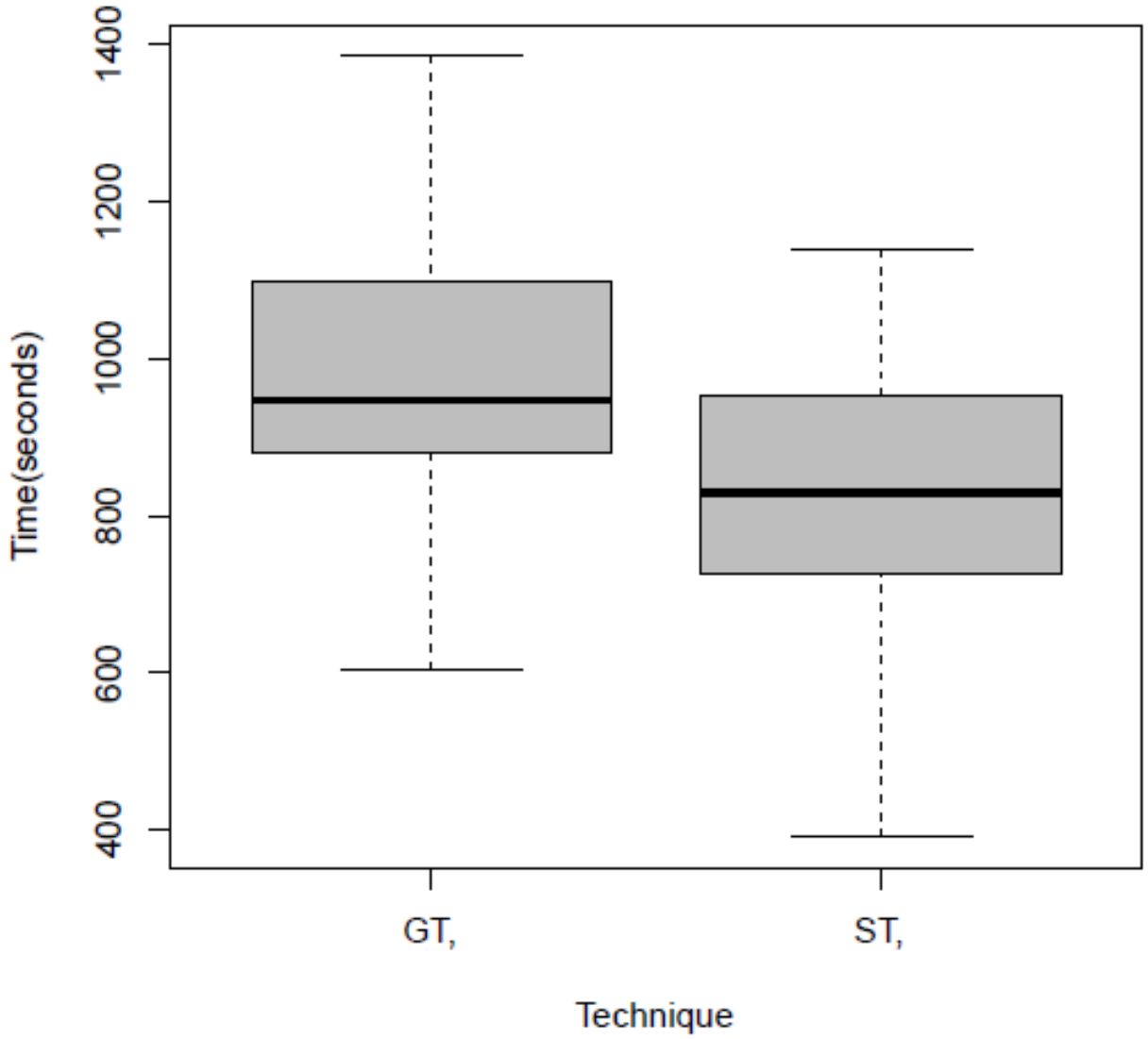
20 (18) UFPE graduate students

RGMS



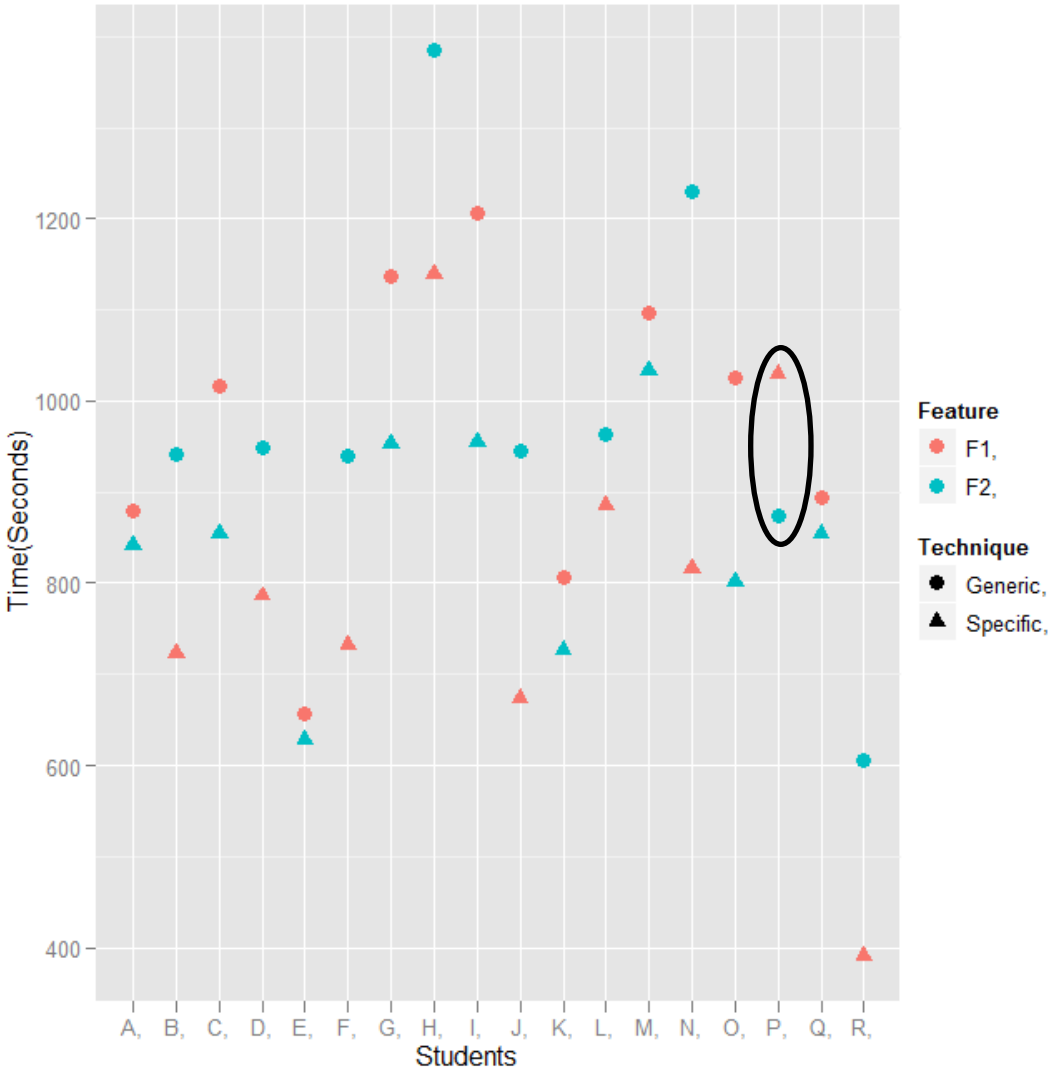
TestWatcher

The ST had better results than the GT



	GT	ST
Average	975	824
Standard Deviation	192	172

Fourth experiment individual results



ANOVA's p-value for technique factor: 0.0001082

Terminated CRs

	Valid	**Invalid**
ST	18	1
GT	15	20

Threats to internal validity

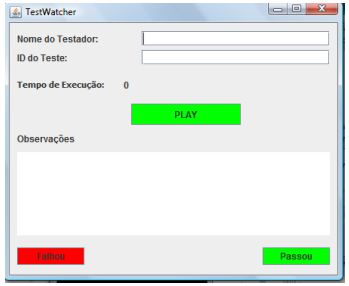
- Configuration of Latin square replicas
  - Raffle the techniques for each replica
- Heterogeneous environment

FIFTH EXPERIMENT

Tools and participants

22 (10) UNB undergraduate students

RGMS

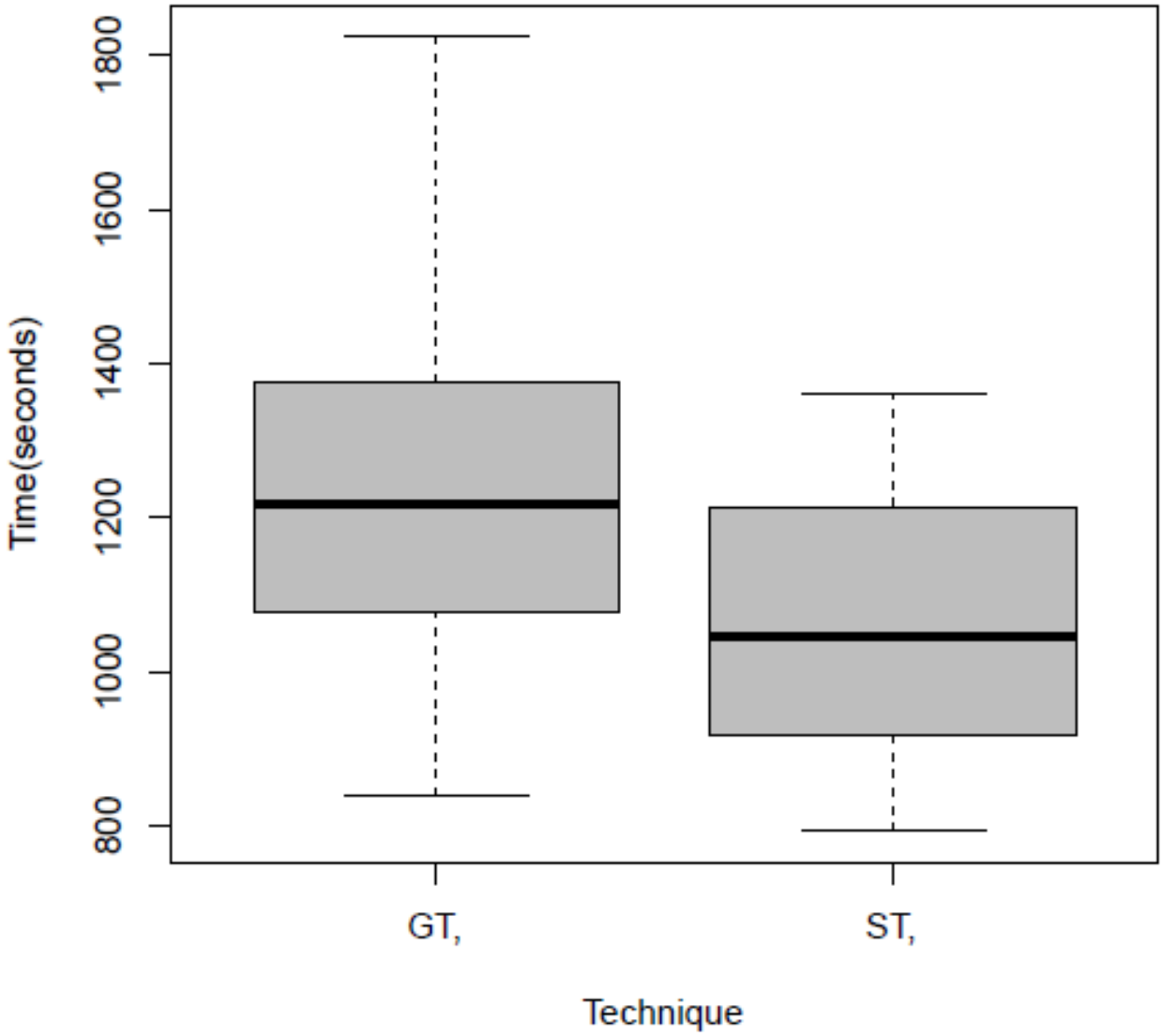


TestWatcher

Time collection approach

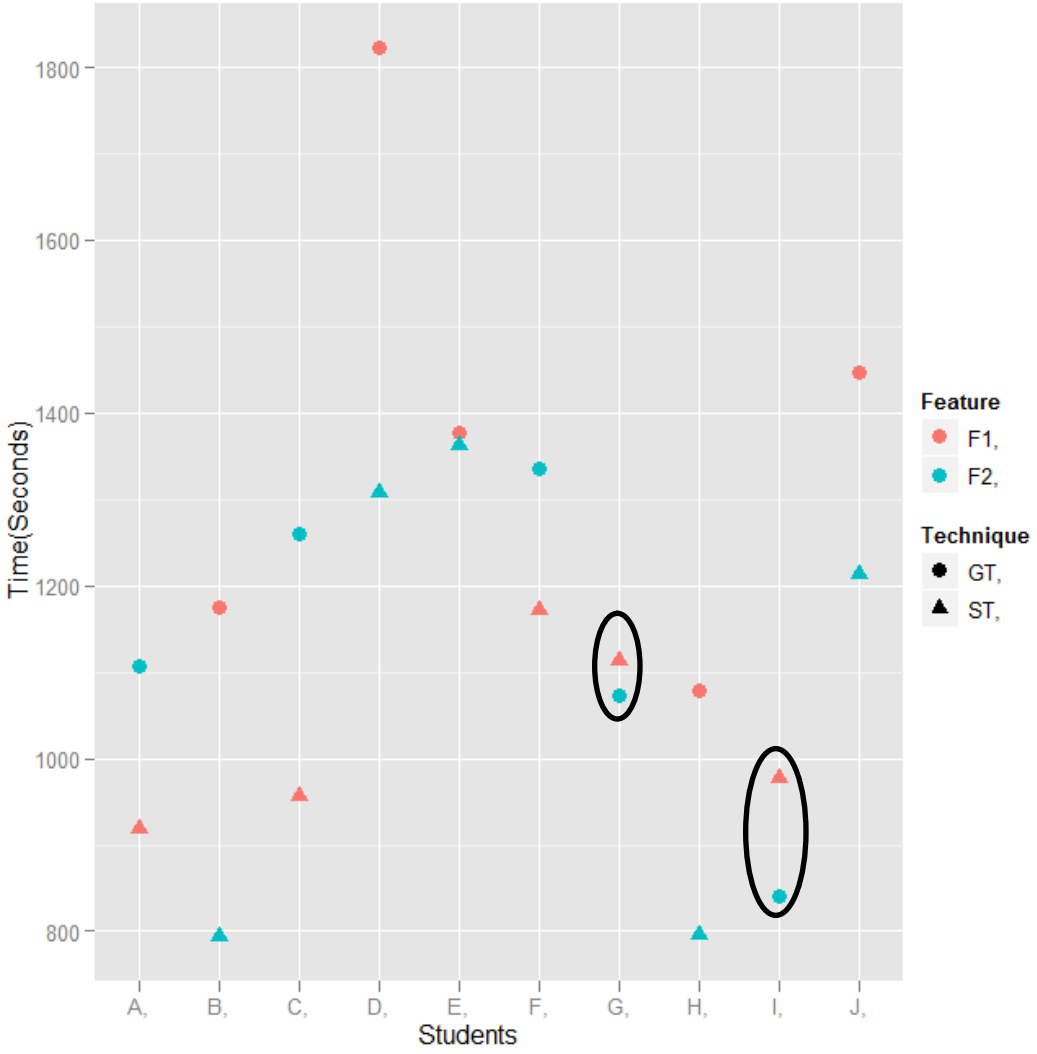
Collecting execution time together with CR report time

Fifth experiment box-plot



	GT	ST
Average	1251	1061
Standard Deviation	268	203

Fifth experiment individual results



ANOVA's p-value for technique factor: 0.01085

Terminated CRs

	Valid	**Invalid**
ST	9	1
GT	13	20

Threats to internal validity

- Size of reported CRs can vary from one participant to the other
- Fix test cases entry values

General threats to internal validity

- RGMS not being an SPL from the real world
- Benefits [Buse, 2011]:
  - Less training
  - Simplify recruiting
  - Greater control over confounding factors

General threats to external validity

- Using students as participants [Staron, 2007]
- Different product lines can benefit in different ways from specific test cases
- Testers with different experience in one determined SPL may benefit differently

Future work

- Systematic mapping study to bring up existing techniques that support functional test cases development for SPL
- Evaluate existing techniques using empirical methods
- Improve TaRGeT to incorporate SPL variability constructs

Comparing Different Test Strategies for Software Product Lines

Paola Accioly (prga@cin.ufpe.br)

Advisor: Paulo Borba (phmb@cin.ufpe.br)
Co-advisor: Rodrigo Bonifácio (rba2@cin.ufpe.br)