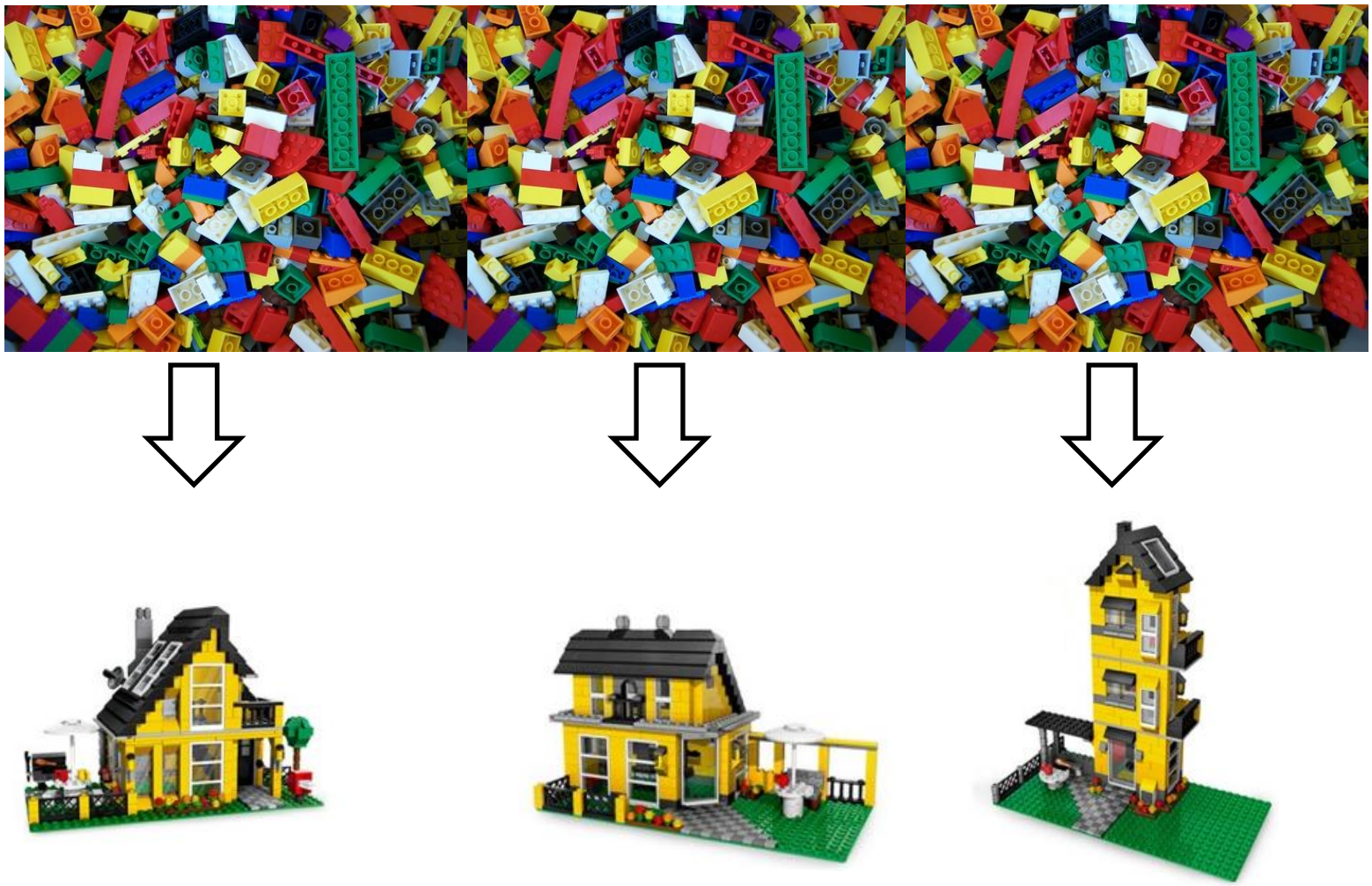
Comparing Different Test Strategies for Software Product Lines

Paola Accioly (prga@cin.ufpe.br)

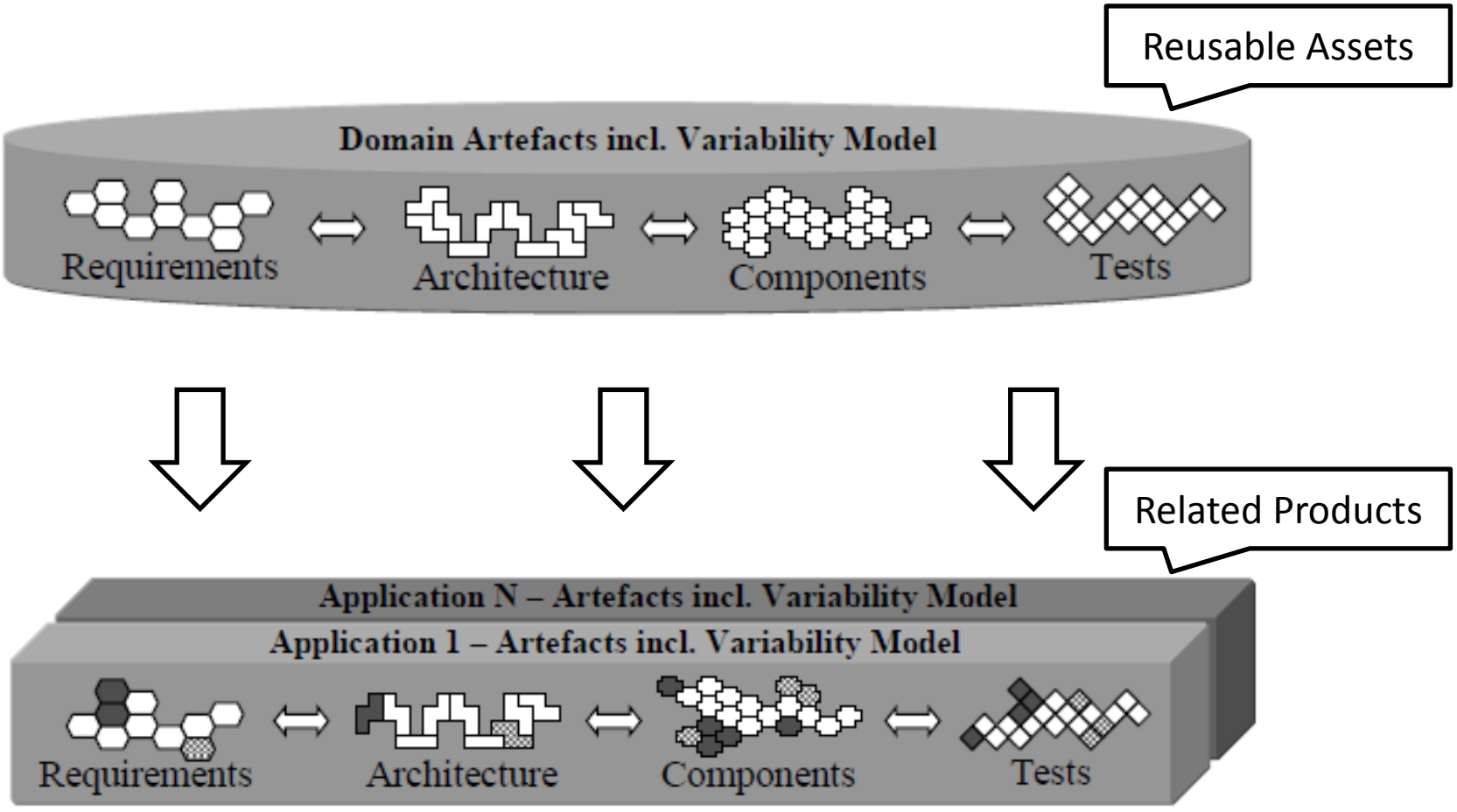
Advisor: Paulo Borba (phmb@cin.ufpe.br)
Co-advisor: Rodrigo Bonifácio (rba2@cin.ufpe.br)

CONTEXT

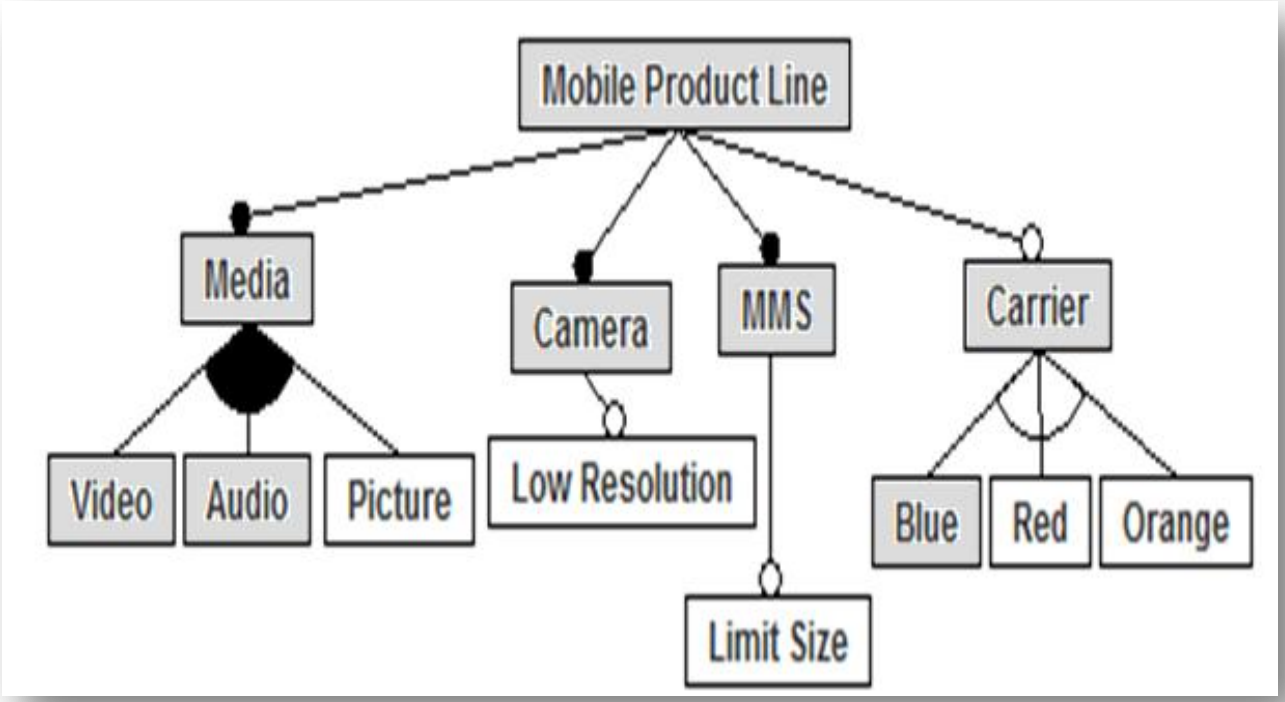
Software product lines



Domain vs. application

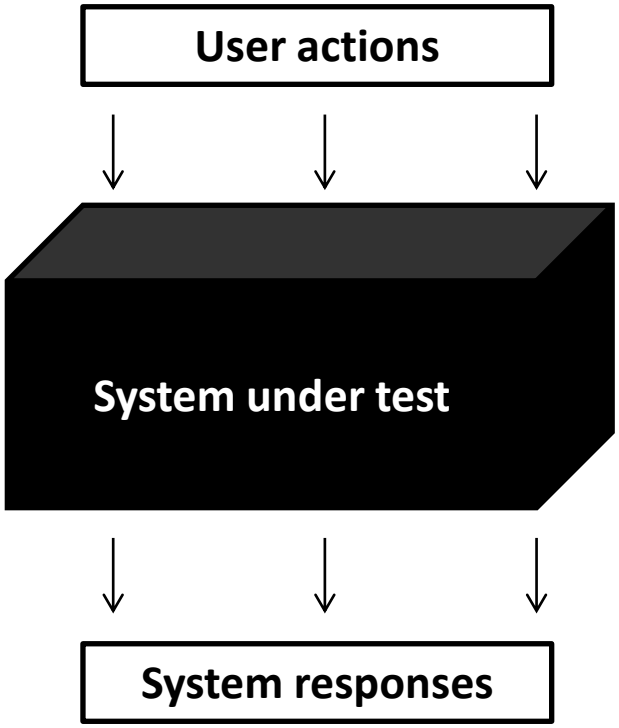


Feature model



Product 1:
Mobile Product Line
Media: Video, Audio
Camera
MMS
Carrier: Blue

Black-box testing strategy

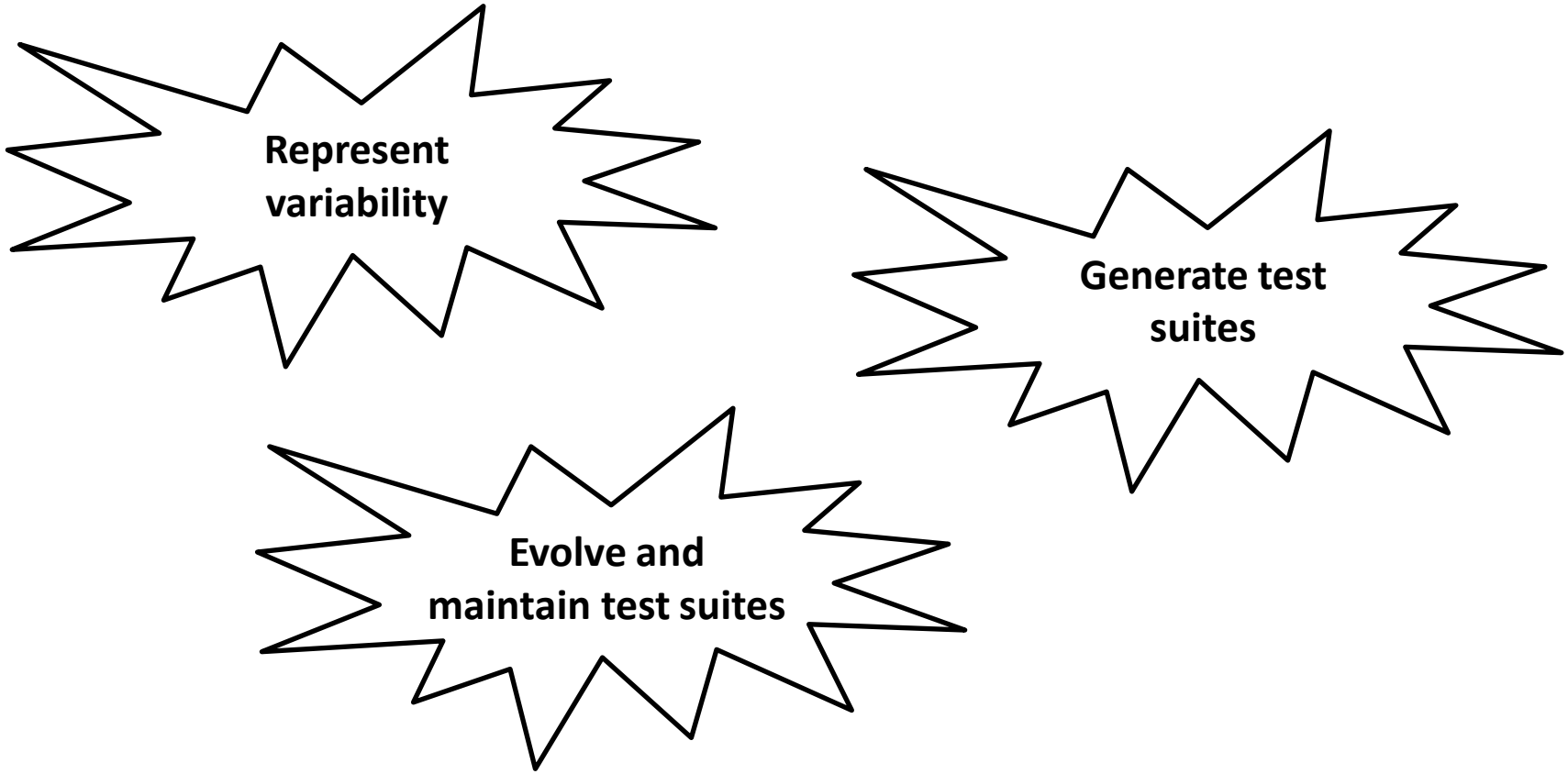


Specification based test cases

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’	Message is correctly sent

MOTIVATION

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

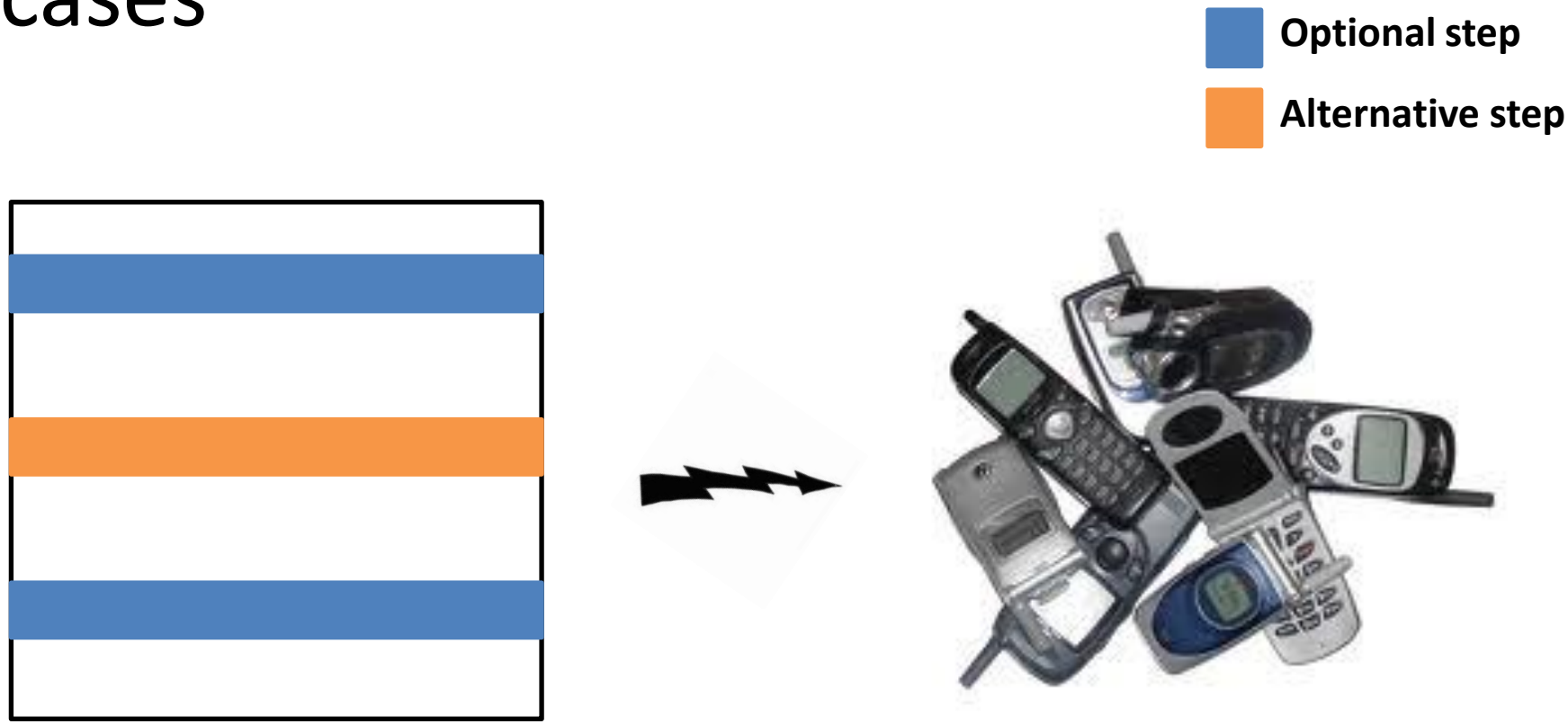


Recently some techniques have been proposed…

However, they still require further evaluation.

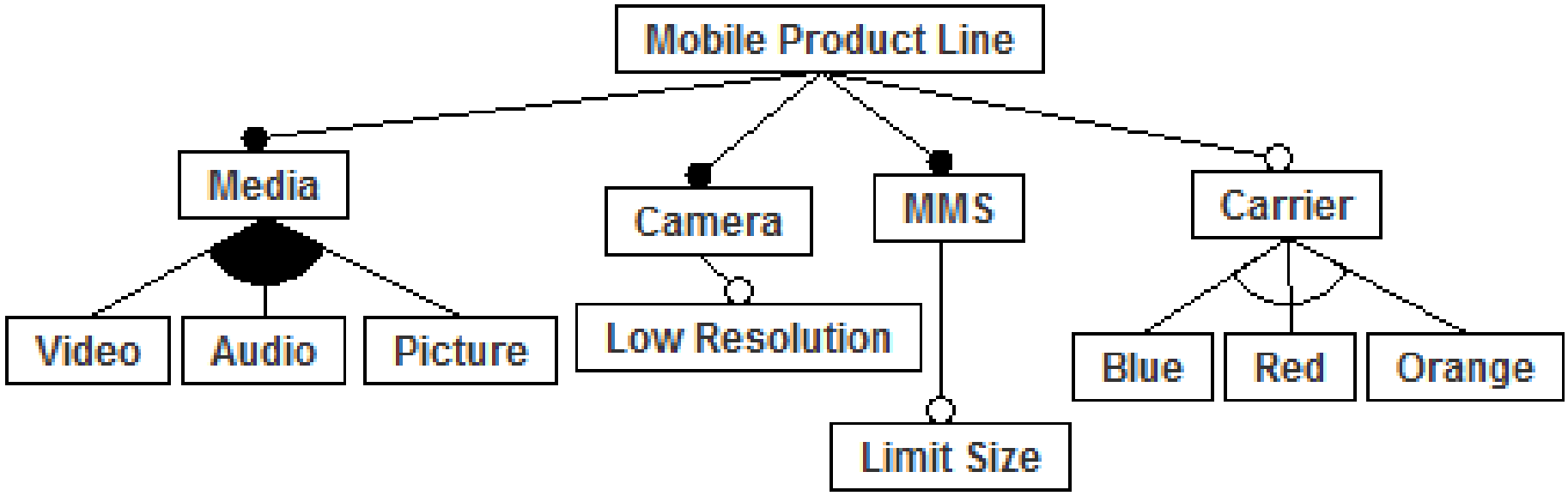
Consequently, the industry is not encouraged to invest in adopting such techniques.

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 some examples!

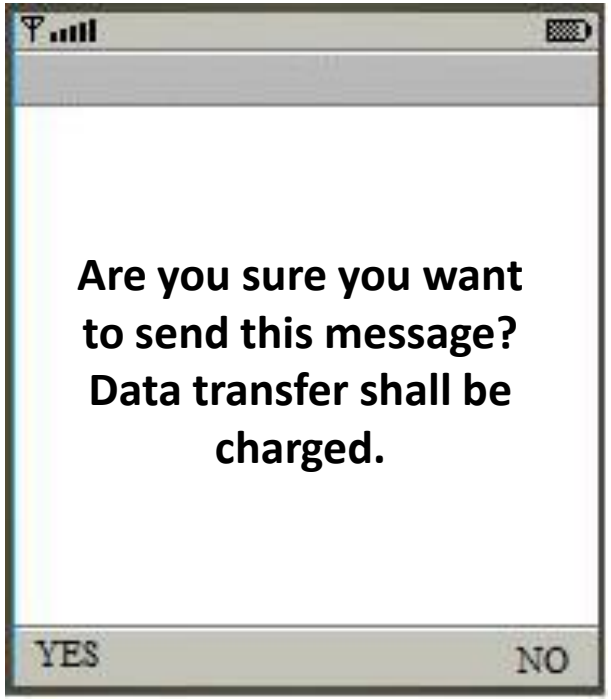
Mobile product line



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

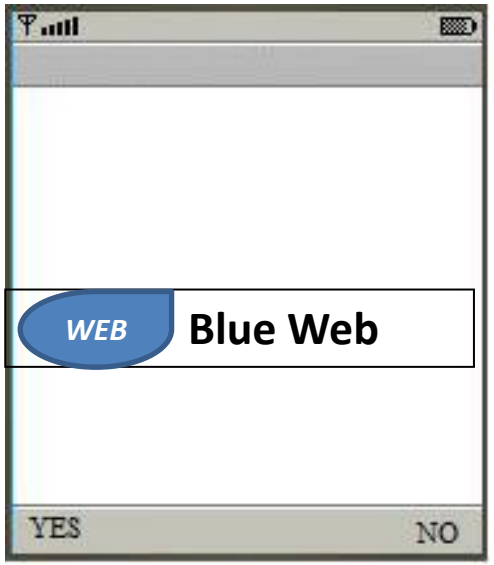
Specific test case for products configured with the Blue carrier feature

Step 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’	**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**

Test case: user checks icon and label on mobile main menu

Step nº	User Action	System Response
1	Go to Main Menu	Main Menu Appears
2	See that there is an option with this icon called “Web”	Icon and Title appears correctly

Test Case



Product Behavior

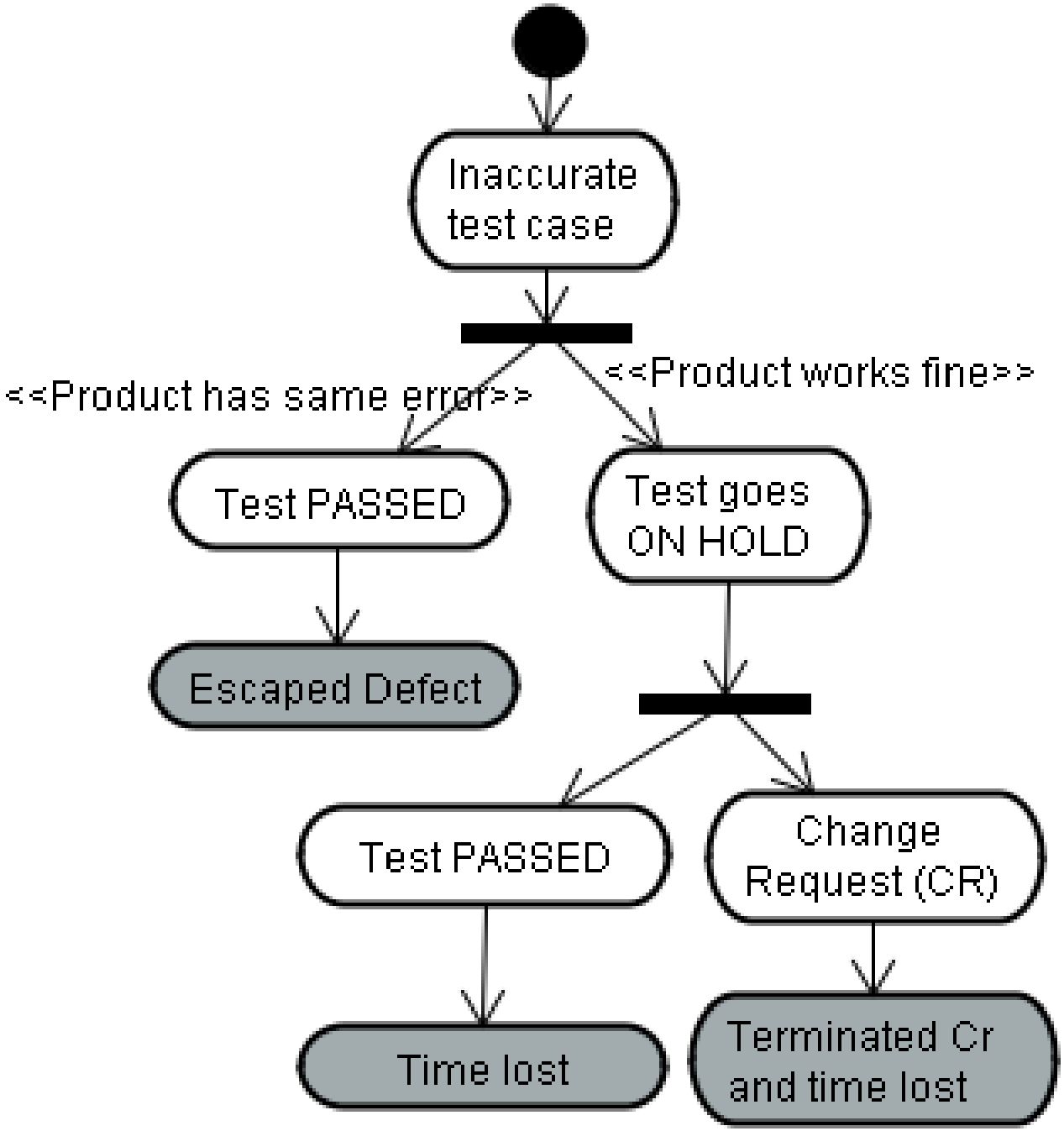
Test Case: User Attaches Video to MMS

Step №	User Action	System Response
1	Go to Main Menu	Main Menu appears
2	Open Camera Application	Camera App opens
3	Make a 5s video	Video is correctly saved to phone memory
4	Select Options	Option Menu appears
5	Select “Send as MMS”	Dialog appears: “Video is too large to attach. Do you want to resize it?”
6	Hit “Yes”	Video is correctly resized and attached
7	Add recipient	Recipient is added
8	Select “Send Message”	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 kind 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

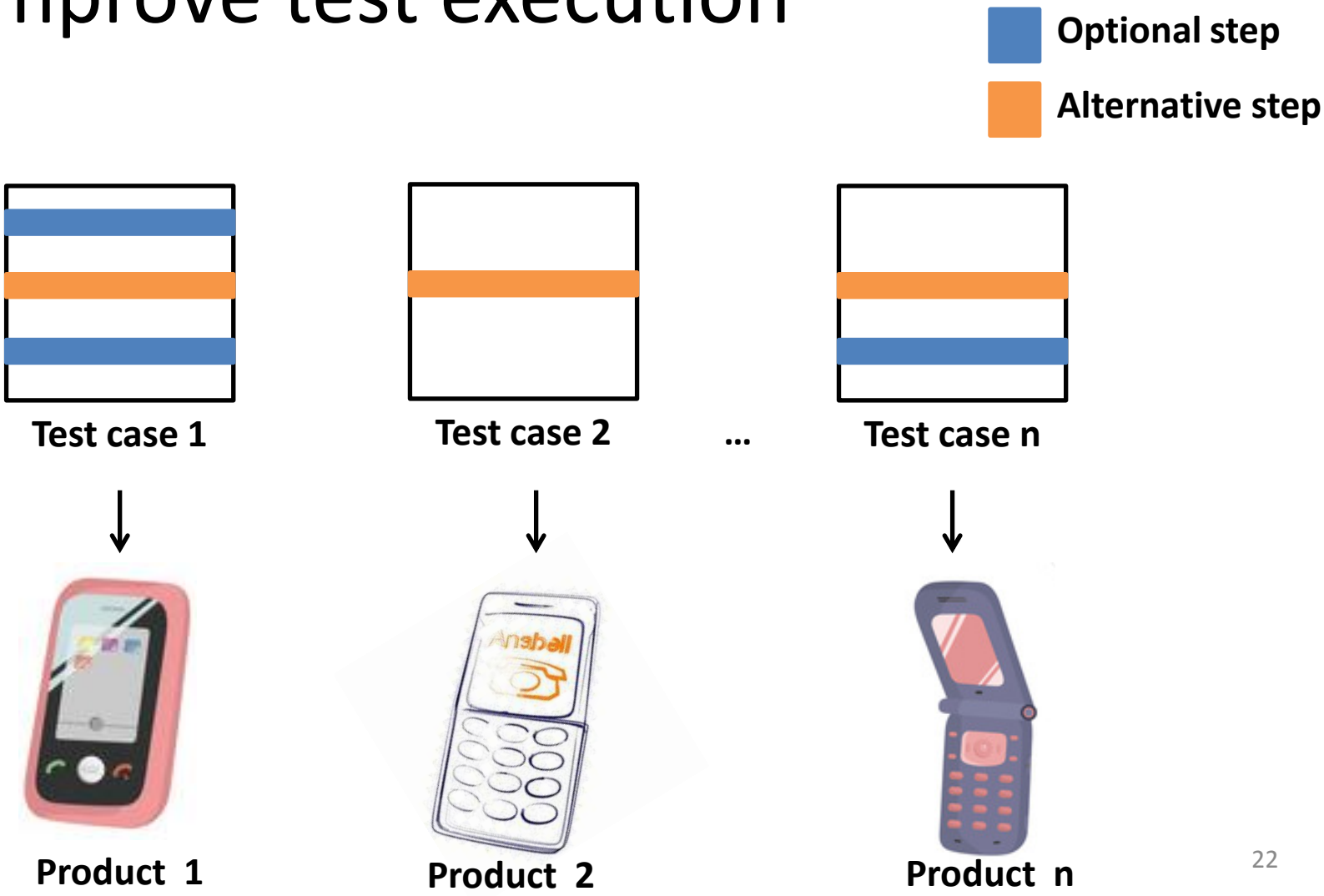
Escaped Defects affect directly the products quality

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

Error prone

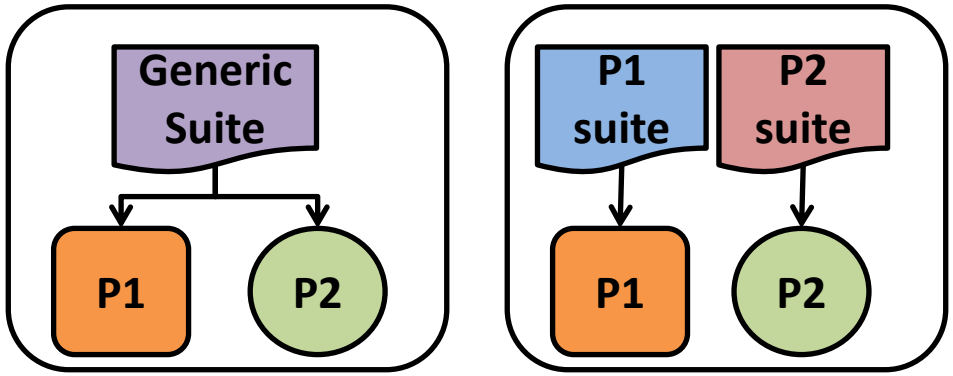
Low productivity

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



Our proposal

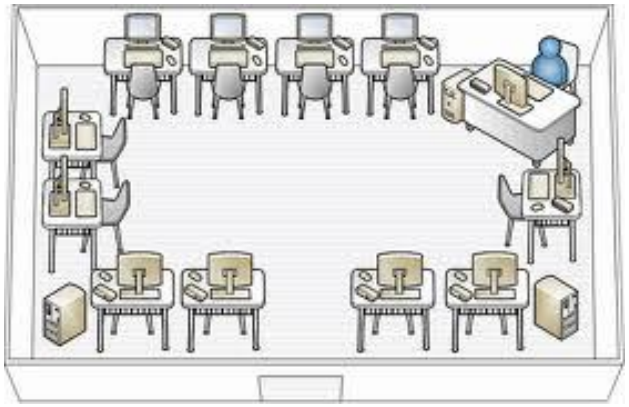
To compare both techniques (generic vs. specific) to investigate their impact from the point of view of the test execution process



Empirical software engineering

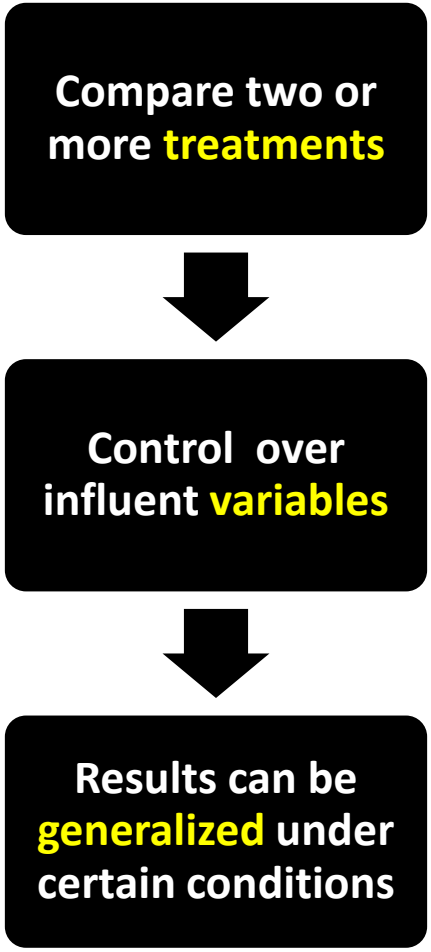
Survey

Case study



Controlled experiment

Controlled experiments



EVALUATION STUDIES

GQM

Goal:

Analyze test execution metrics, for the purpose of evaluating the effect of two different test case design techniques for SPL (GT vs. ST), with respect to their effectiveness regarding time to execute the test suites and the number of terminated CRs reported during the test execution process. Using the point of view of test engineers and software engineering researchers in the context of experiments done with software engineering students in the environment of universities.

Questions

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

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

Metrics

Test execution time

Number of Terminated CRs

The Latin square design

	Feature1	Feature2
Subject1	**GS**	**ES**
Subject2	**ES**	**GS**

	Feature1	Feature2
Subject3	**ES**	**GS**
Subject4	**GS**	**ES**

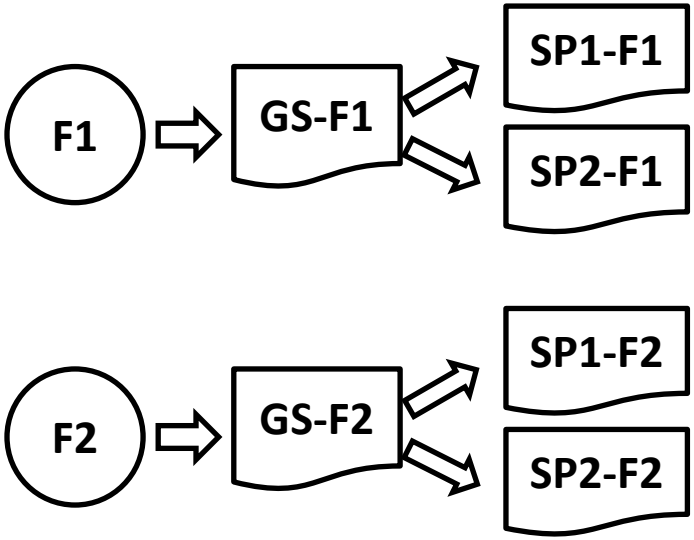
...

GS– Generic Suite
ES– Specific Suite

Experiment Operation

DAY 1	DAY 2	DAY 3
Training and dry-run	Latin square first round	Latin square second round

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.

FIRST EXPERIMENT

Tools and participants

Use Case Specifications

Test Suites

Participants: 7 computer science undergraduate students from UFPE

Manual collection of time

Threats to internal validity

- Manual time collection
- Time collected as a whole(Setup + Execution + Debug)
- TaRGeT
- Feature had test cases with similar steps

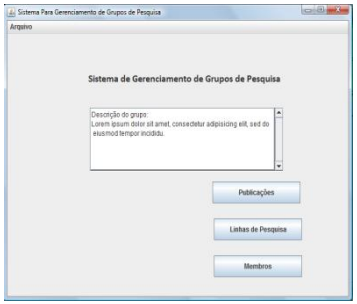
SECOND EXPERIMENT

Tools and participants

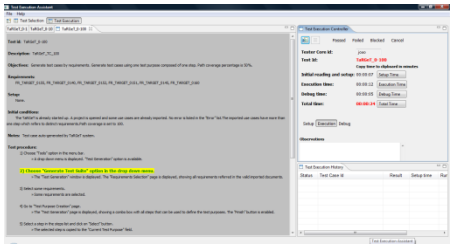
6 UFPE graduate students

+

3 UFPE undergraduate students

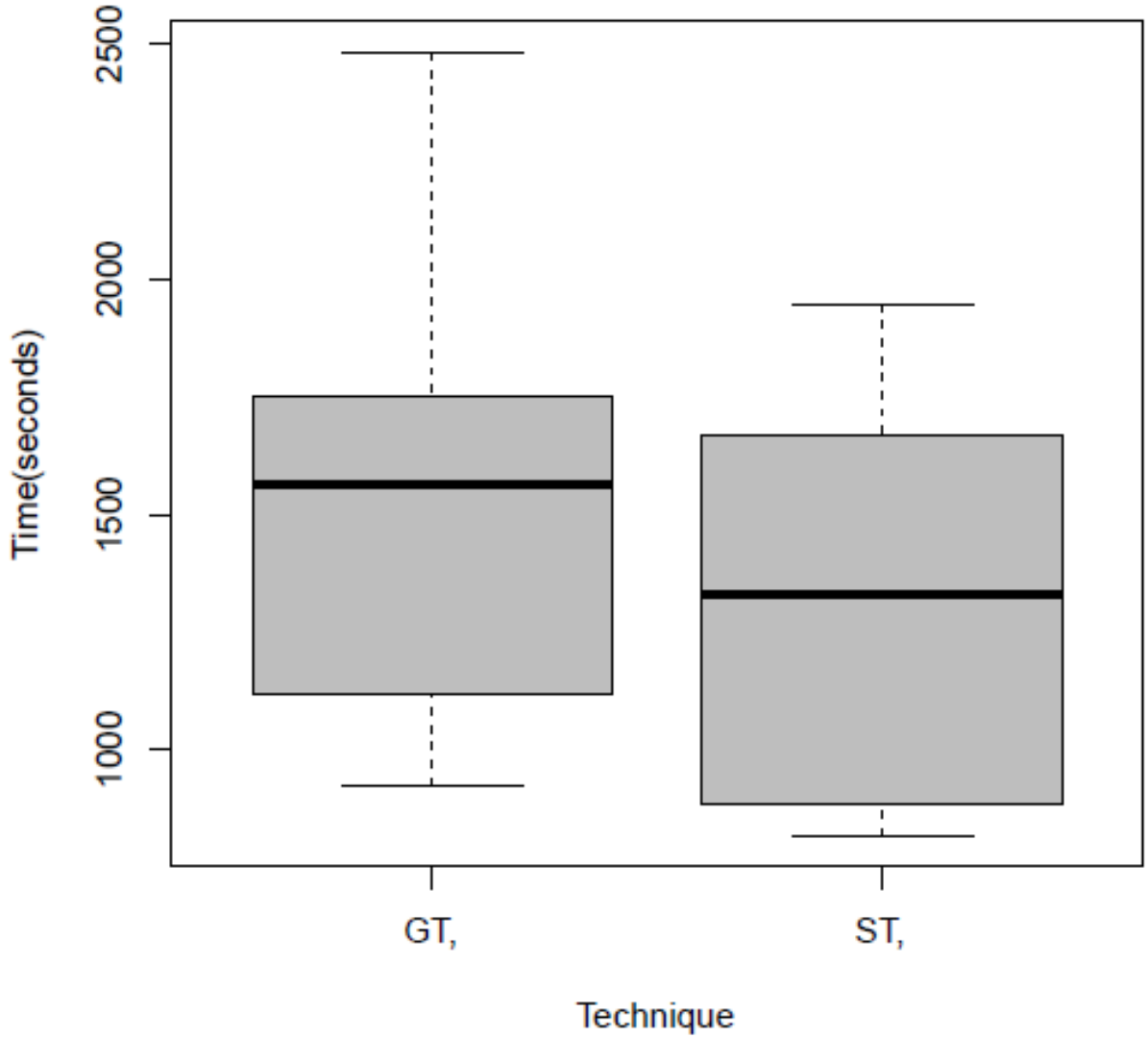


Research Group Management System (RGMS)

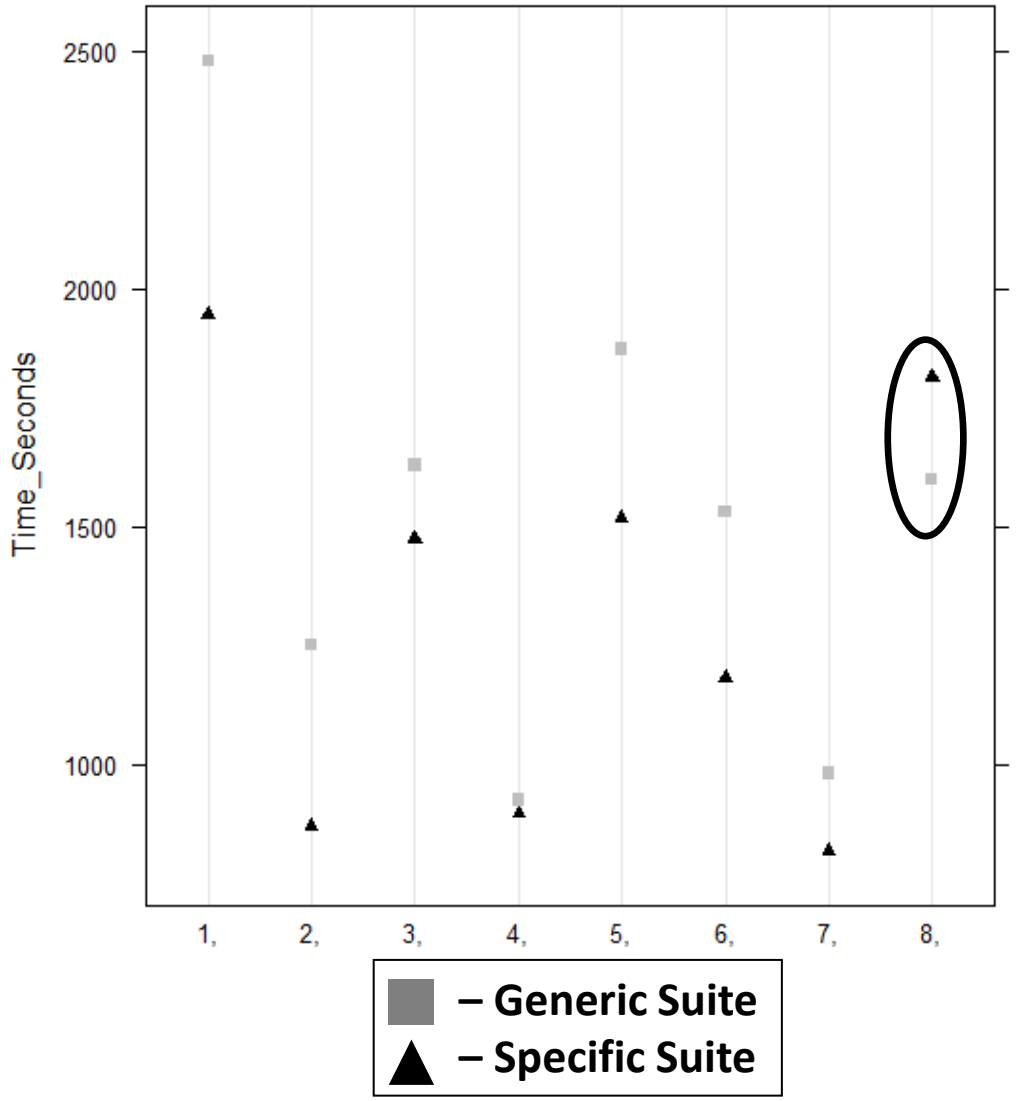


ManualTEST

Box-plot



Individual results



ANOVA

	Df	Sum Sq	Mean Sq	F value	*p-value*
Replica	3	431280	143760	5.0998	**0.0434014**
Replica:Student	4	2506626	626657	22.2304	**0.0009623**
Feature	1	27556	27556	0.9775	0.3609867
Technique	1	191844	191844	6.8056	**0.0401876**
Residuals	6	169135	28189		

Reported CRs

	Valid	Invalid
ST	12	0
GT	12	2

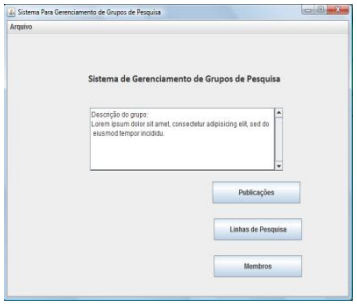
Threats to internal validity

- ManualTEST
- Time collection approach

THIRD EXPERIMENT

Tools and participants

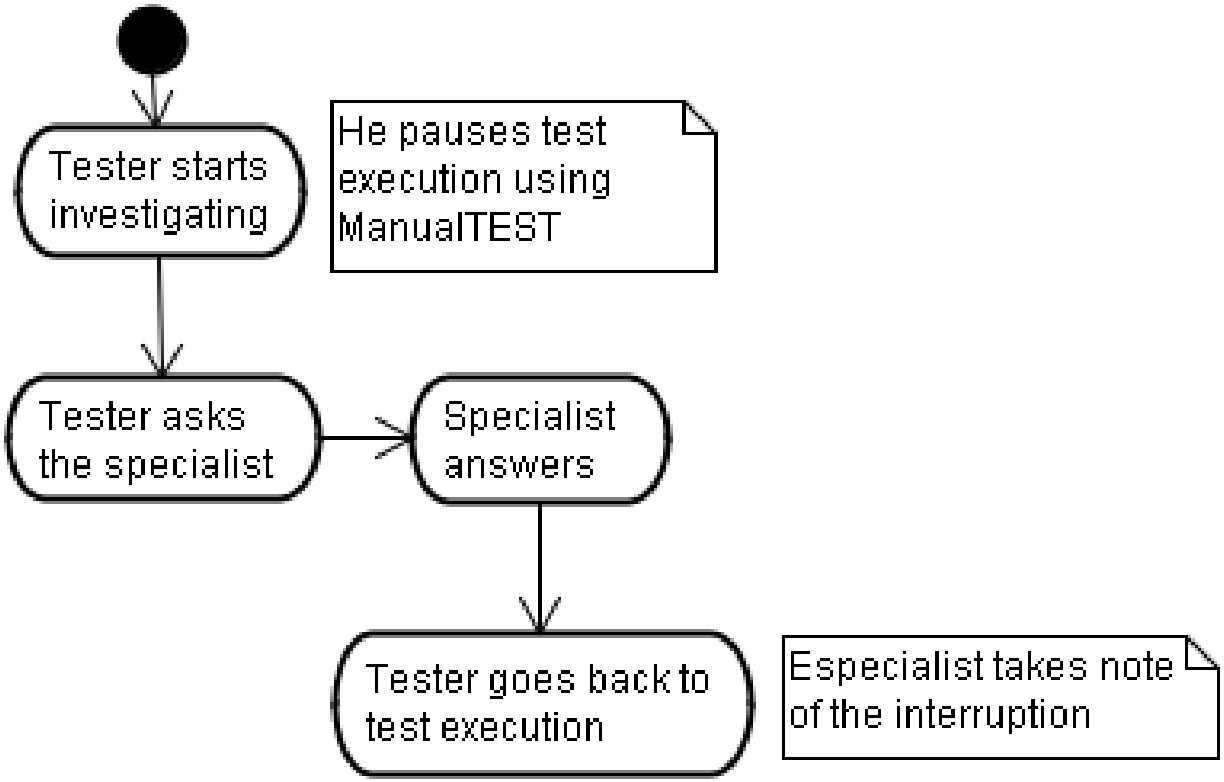
20 UNB (University of Brasília) undergraduate students



RGMS

TestWatcher

Time collection approach



Threats to internal validity

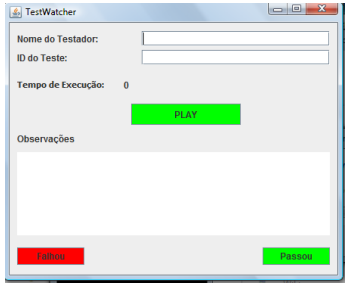
- Lack of a dry-run
- Low attendance to class

FOURTH EXPERIMENT

Tools and participants

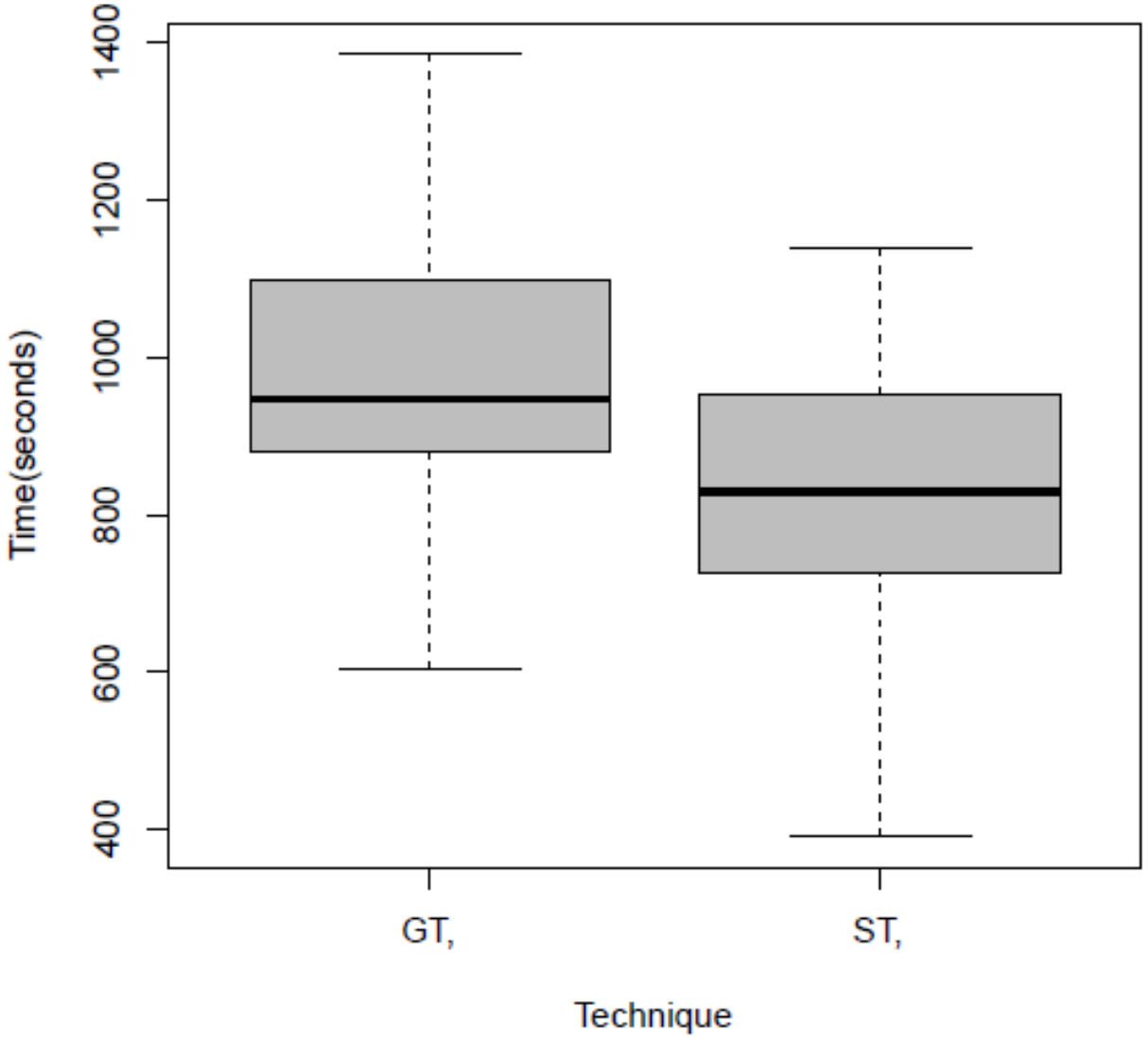
20 UFPE graduate students

RGMS

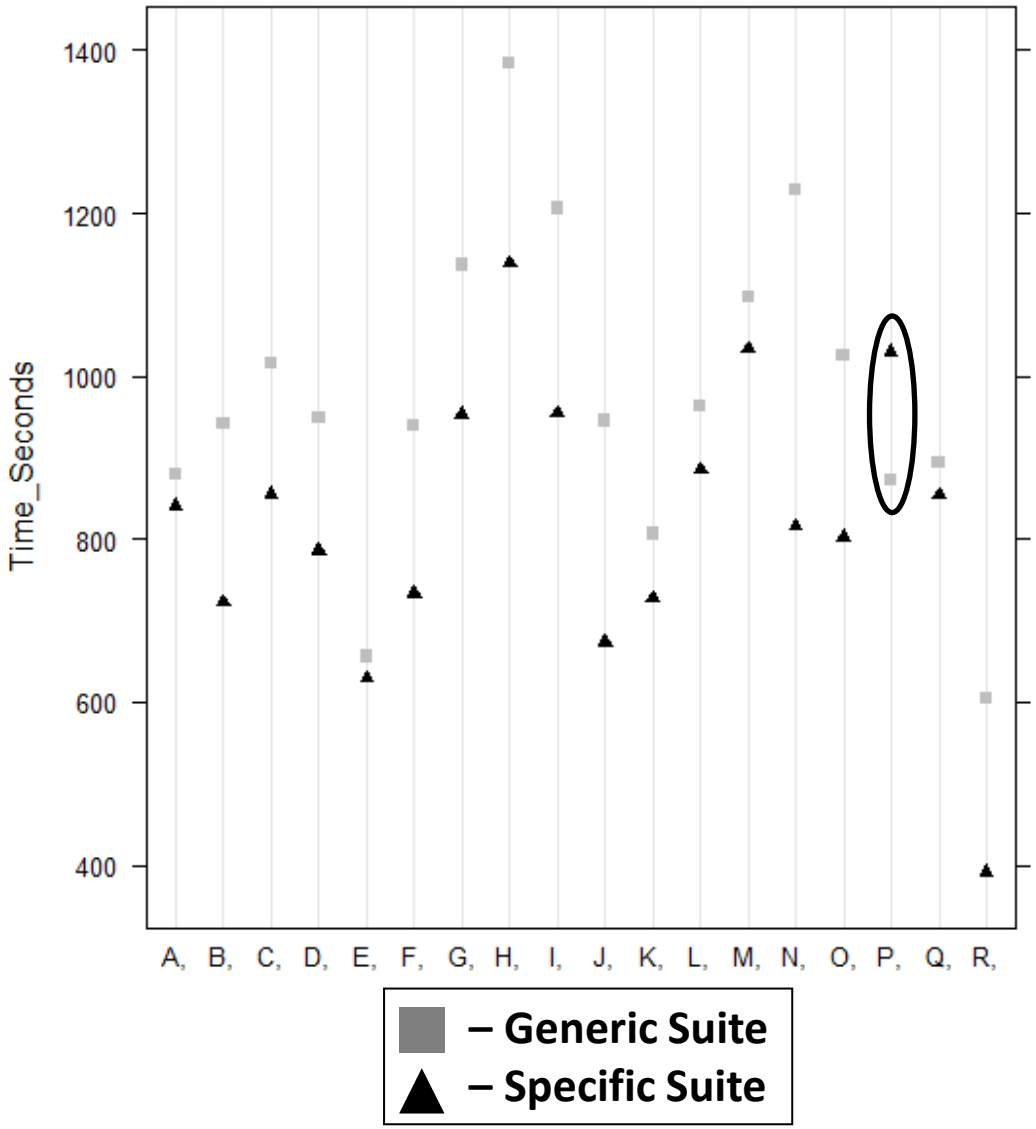


TestWatcher

Box-plot



Individual results



ANOVA

	Df	Sum Sq	Mean Sq	F value	*p-value*
Replica	8	661378	82672	10.4027	**4.748e-05**
Replica:Student	9	333278	37031	4.6596	**0.0037052**
Feature	1	9571	9571	1.2044	0.2886915
Technique	1	206267	206267	25.9548	**0.0001082**
Residuals	16	127155	7947		

Terminated CRs

	Valid	Invalid
ST	18	1
GT	15	20

Threats to internal validity

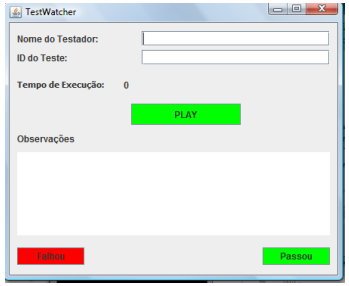
- Configuration of Latin square replicas
- Heterogeneous environment

FIFTH EXPERIMENT

Tools and participants

22 UNB undergraduate students

RGMS

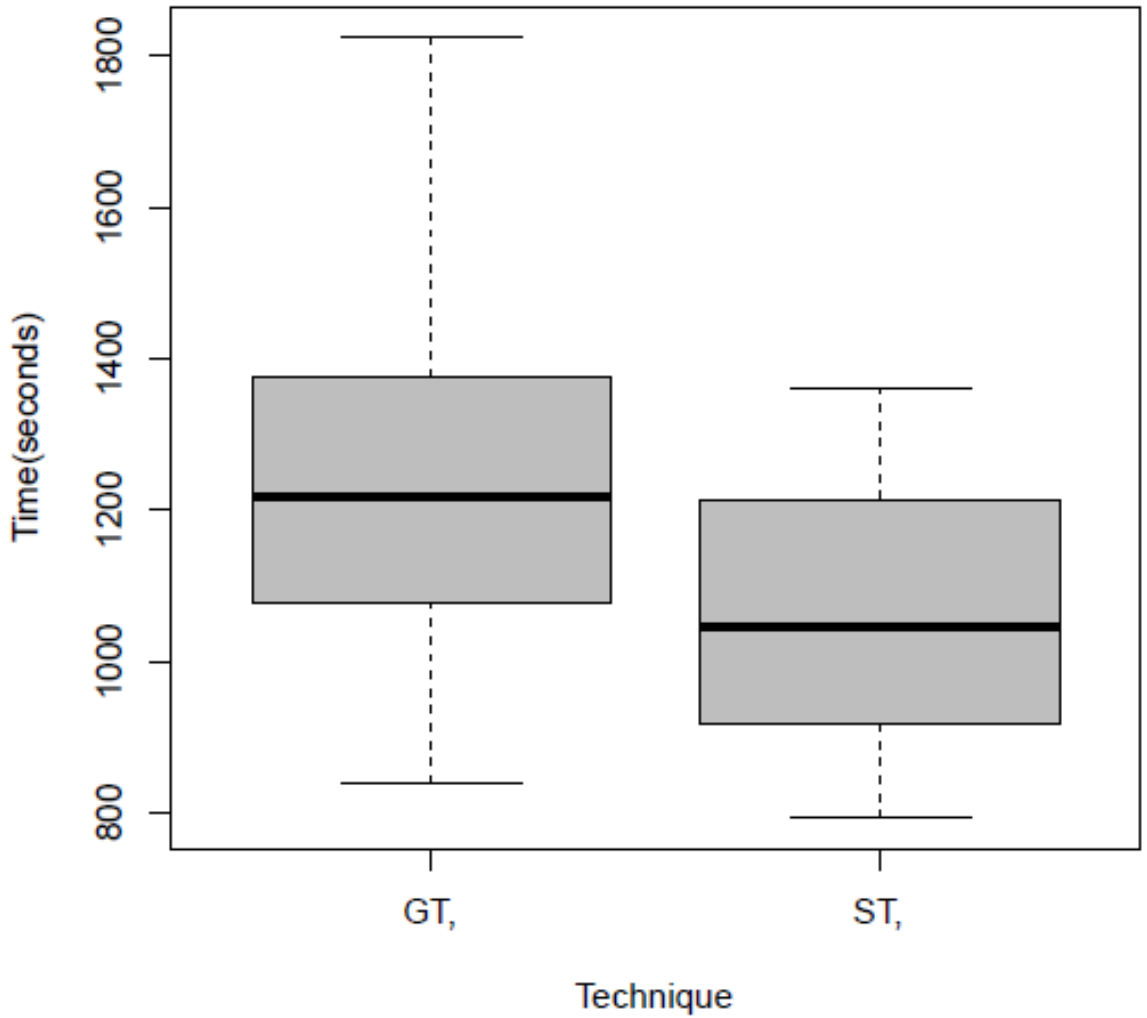


TestWatcher

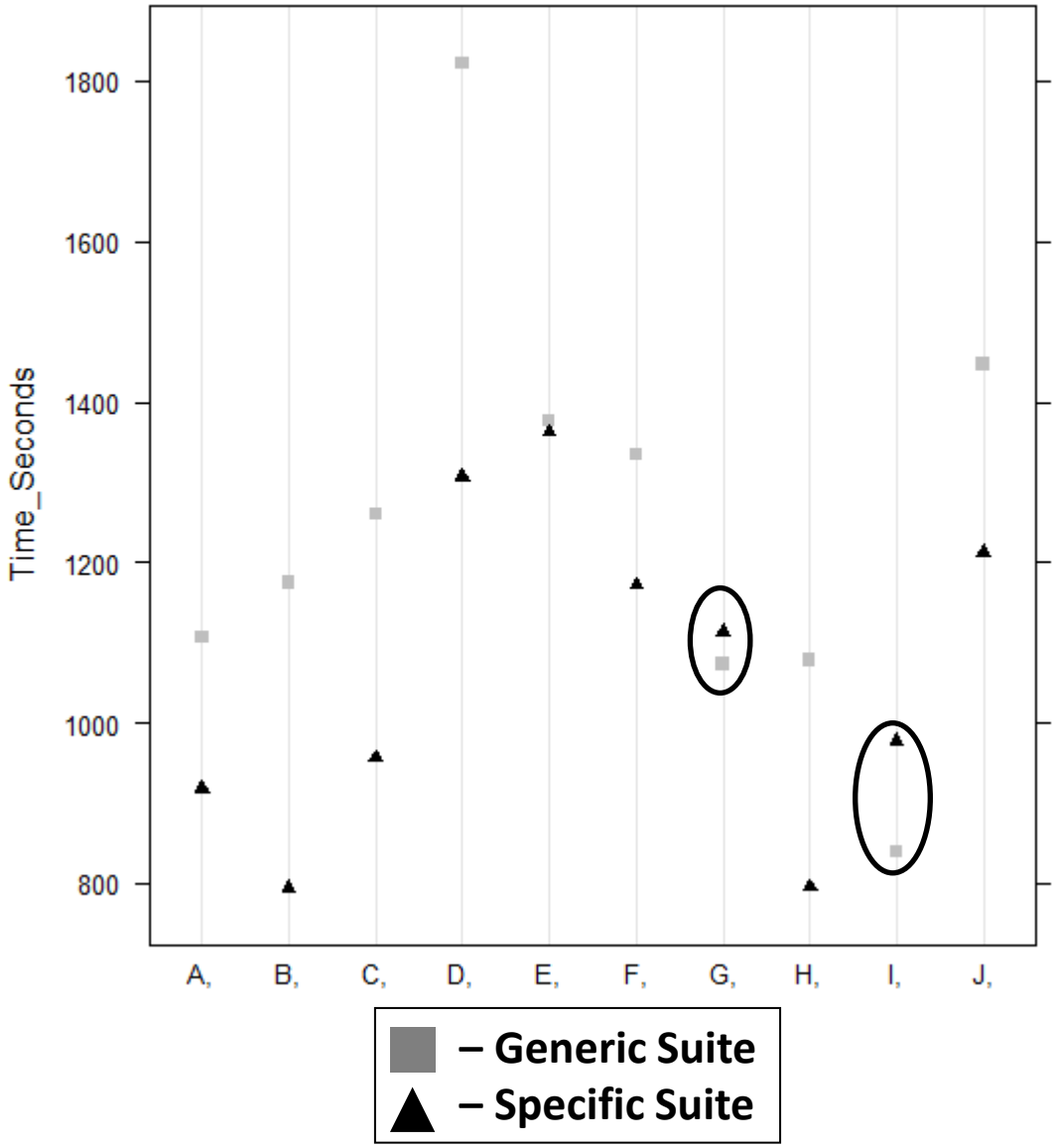
Time collection approach

Collecting execution time together with CR report time

Box-plot



Individual results



ANOVA

	Df	Sum Sq	Mean Sq	F value	*p-value*
Replica	4	411791	102948	6.1673	**0.01446**
Replica:Student	5	425216	85043	5.0947	**0.02143**
Feature	1	45220	45220	2.7090	0.13840
Technique	1	181832	181832	10.8931	**0.01085**
Residuals	8	133540	16693		

Terminated CRs

	Valid	Invalid
ST	20	1
GT	13	9

Threats to internal validity

- Size of reported CRs
- RGMS

Threats to external validity

- Using students as participants
- Different product lines can benefit in different ways from specific test cases

CONCLUDING REMARKS

Summary

- Although some techniques to specify black box test cases were proposed, the research community still lacks of empirical evaluations
- Industries do not invest in adopting those techniques
- We have executed 5 controlled experiments to evaluate the effect of generic and specific test cases from the point of view of the test execution process
- 3 experiments (2nd, 3rd and 5th) gathered evidence that specific test cases can increase productivity in the test execution environment

Related work

Techniques to manage functional test cases for SPL

Bertolino and Stefania Gnesi
Nebut et al.
Pohl et al.

Empirical studies on SPL testing

Ivan et al.
Ganesan et al
Denger and Kolb

Empirical studies on software testing

Itkonen et al
Lima et al

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 MSVCM constructs

Comparing Different Test Strategies for Software Product Lines

Paola Accioly (prga@cin.ufpe.br)
Adviser: Paulo Borba (phmb@cin.ufpe.br)
Co-adviser: Rodrigo Bonifácio (rba2@cin.ufpe.br)