# Effectiveness of Exploratory Testing

## An empirical scrutiny of the challenges and factors affecting defect detection efficiency

Khurram Bhatti
www.khurram.info

Ahmad Nauman Ghazi
www.Ghazi.eu

www.bth.se
BLEKINGE INSTITUTE OF TECHNOLOGY

# Agenda

- Paper 1:

Exploratory Testing: A Systematic Literature Review

  - Introduction
  - Research questions
  - Research methods
  - Results

- Paper 2:

An Experiment and Survey on the Effectiveness and Efficiency of Exploratory Testing and Traditional Test Case Based Testing

  - Introduction
  - Research questions
  - Research methods
  - Results

- Conclusions

# Paper 1

# Exploratory Testing:
# A Systematic Literature Review

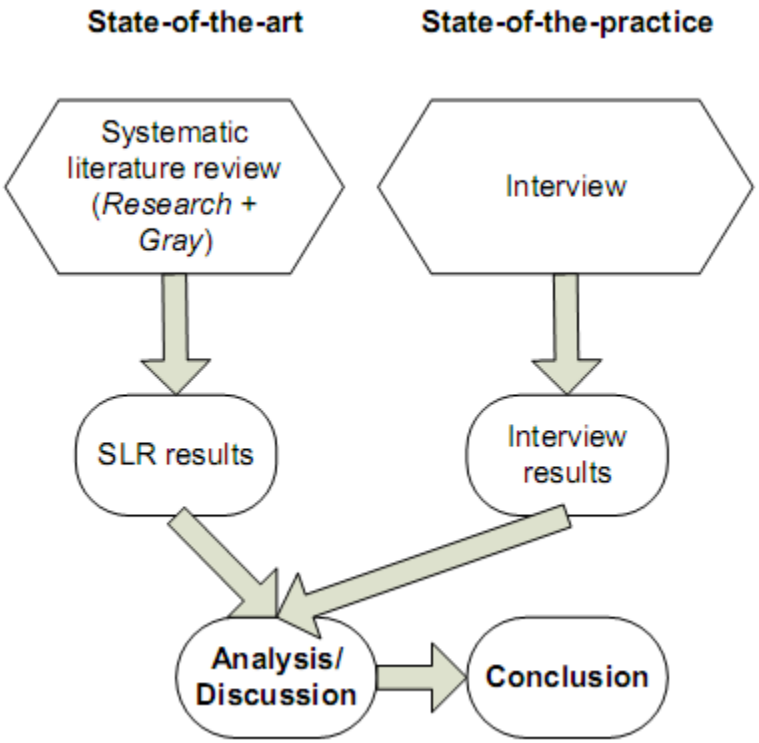
# Research Questions

- **RQ1** What is the state of the art in exploratory testing?
  - **RQ1.1** What definitions of exploratory testing are reported in literature?
  - **RQ1.2** What advantages and disadvantages are reported in literature?
  - **RQ1.3** what challenges are reported in literature?

- **RQ2** What is the state of the practice in exploratory testing?
  - **RQ2.1** How exploratory testing is practiced in testing industry?
  - **RQ2.2** What are the major factors affecting exploratory testing?
  - **RQ2.3** How good exploratory testing approach is for feature coverage?

- **RQ3** What are the challenges and factors in applying exploratory testing? (Findings from RQ1 and RQ2)

# Research methods



# SLR Results

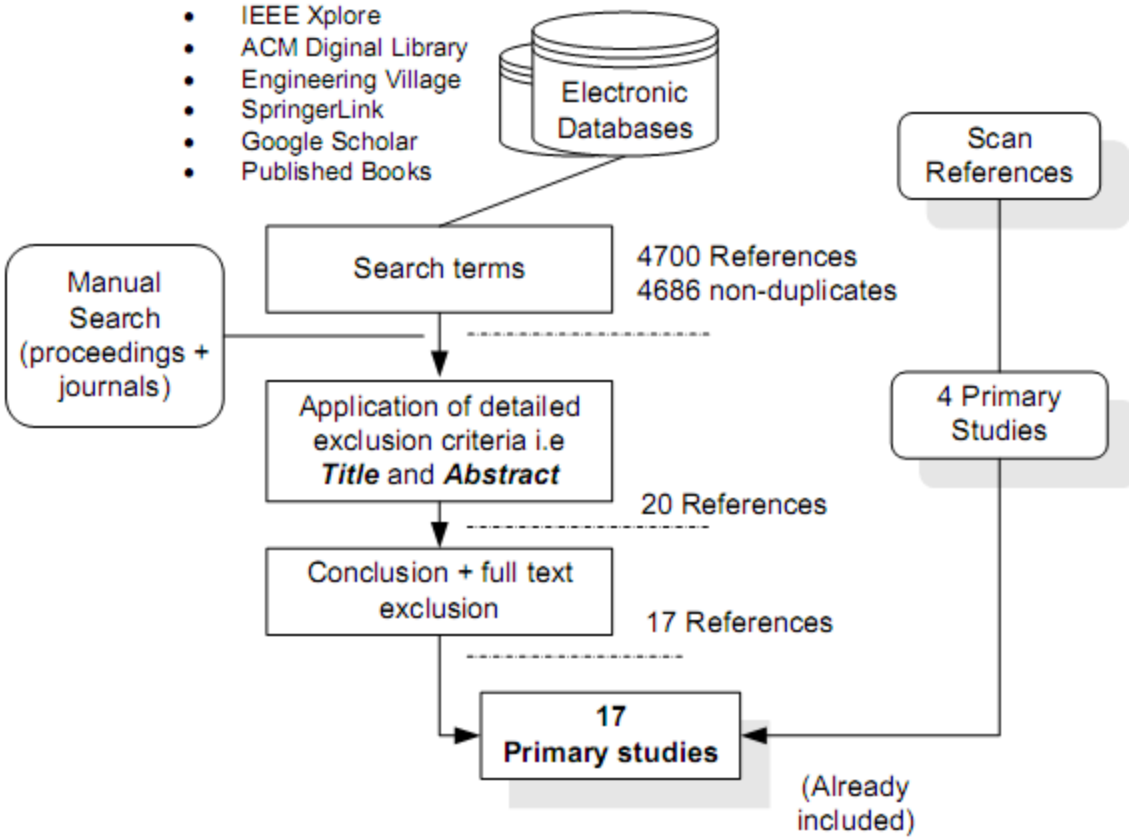
- Primary studies selection

Table 4: Papers before and after duplication removal

| Source | Count | Duplicate | Selected |
|---|---|---|---|
| IEEE Xplore | 171 | 3 | 7 |
| ACM Digital Library | 174 | 2 | 3 |
| Google Scholar | 1,220 | 6 | 2 |
| Engineering village | 1,490 | 5 | 0 |
| SpringerLink | 1,560 | 0 | 0 |
| Published Books | 85 | 3 | 5 |
| **Sub-Total 1** | 4,700 | 14 | 17 |
| Gray Literature(Google) | 58,000 | | 30 |
| **Sub-Total 2** | | | 30 |
| **Total** | | | 47 |

# SLR Results

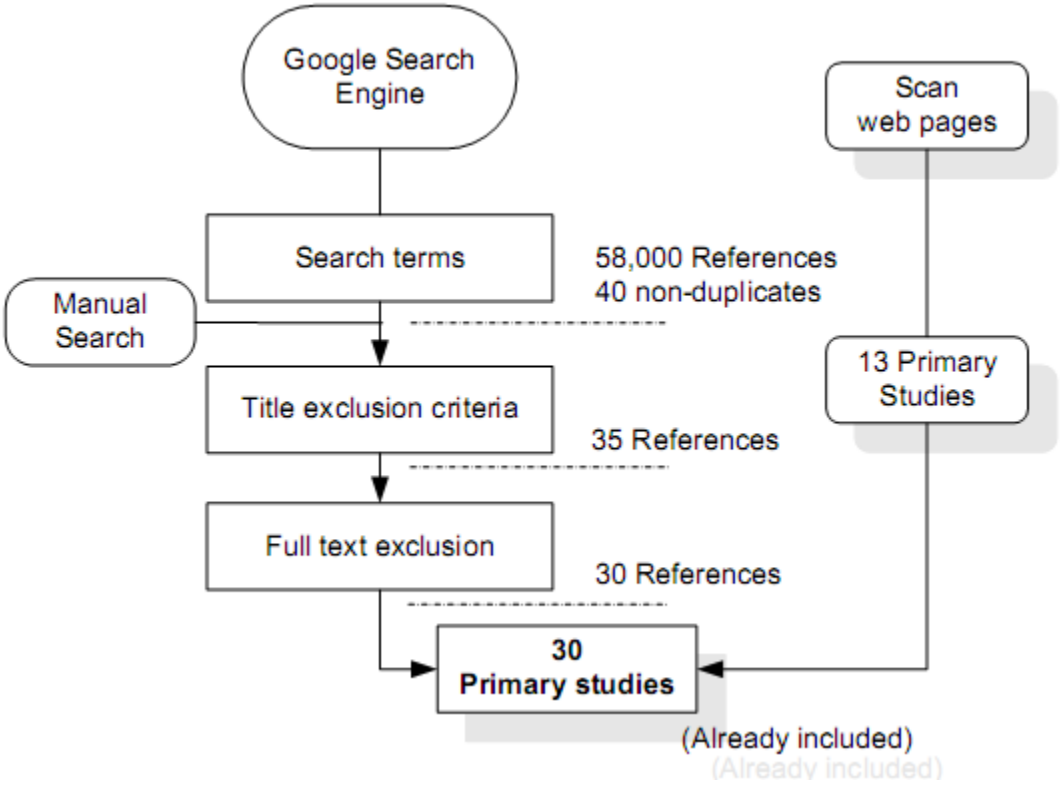
- Research literature




Multi-step filtering of research studies using tollgate approach

# SLR Results

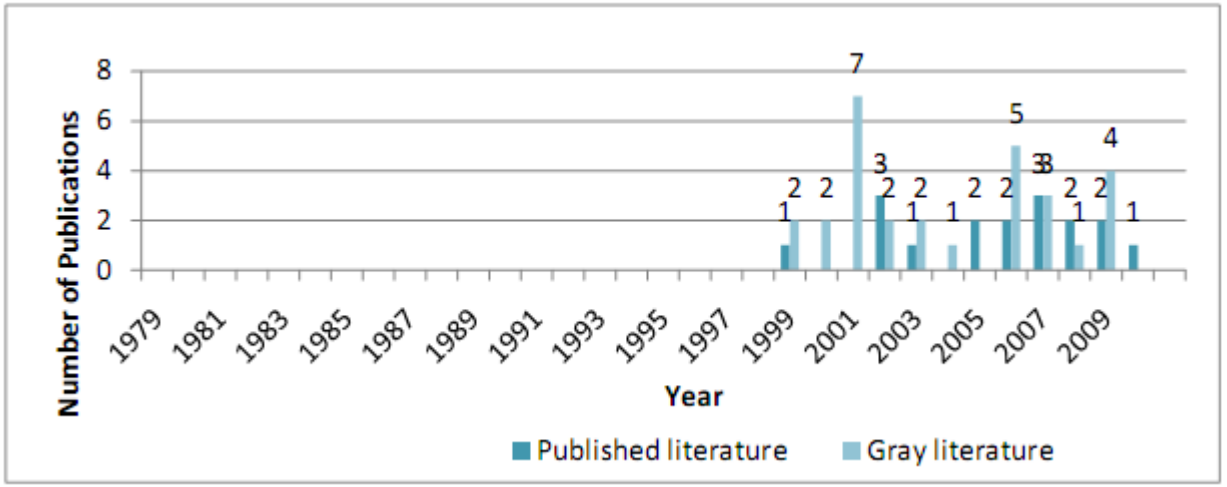
- Gray literature




Multi-step filtering of gray studies using tollgate approach

# SLR Results

- Literature trend




Trend of published and gray literature over the period 1979-2010

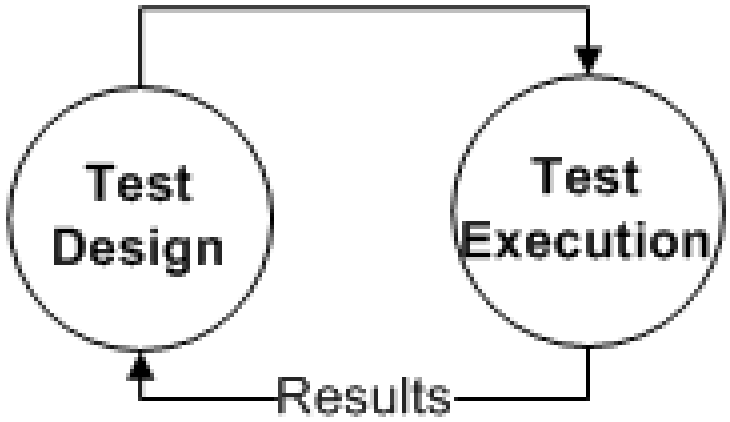
# Exploratory Testing

## Definition

- “Simultaneous learning, test design and test execution”. (James Bach)
- “Any testing to the extent that the tester actively controls the design of the tests as those tests are preformed and uses information gained while testing to design new and better tests”. (Tinkham and Kaner)
- Purposeful wandering, navigating through a space with a general mission, but without a pre-scripted route. Exploration involves continuous learning and experimenting”. (Kaner, Bach and Petticord)

# Exploratory Testing

- No pre-designed test cases
- Controlled by Tester
- Results feedback
- Purposeful wandering




Simultaneous Learning

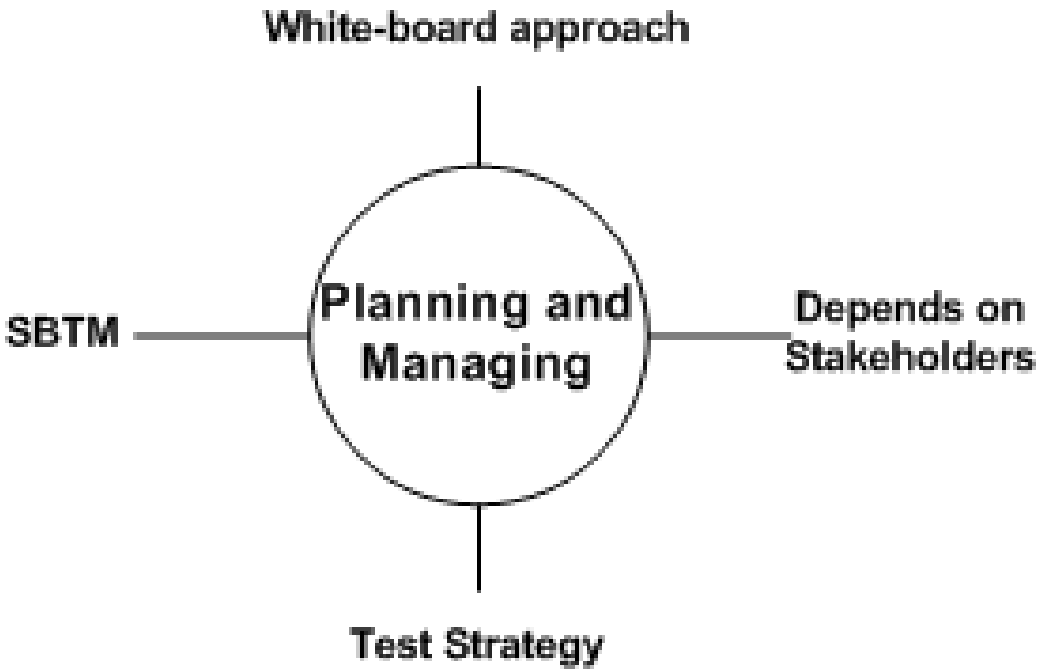
# Exploratory Testing

- Adaptability
  - Learning of system
  - Perform complementary testing
  - Testing under time constraints
  - Targeted testing
- Suitability and non-suitability

# Exploratory Testing

- Challenges
  - Explaining ET to customer and changing their traditional mind set from scripted testing
  - Coaching testers to perform ET professionally
  - Misconceptions related to ET

# Exploratory Testing



# Exploratory Testing

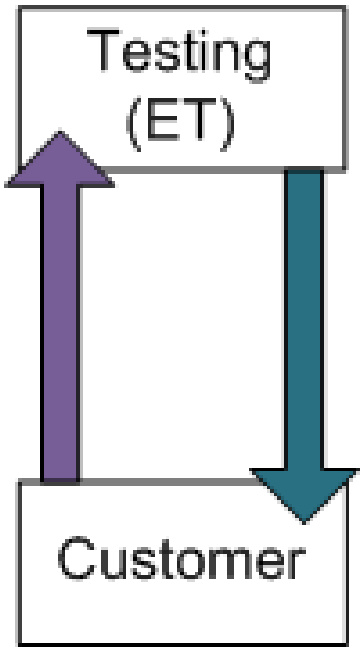
## Factors

- Product
  - Technology
  - Web or client installed
  - Claims
  - Features
  - Customer type
  - User types
- Technology
  - Application
  - Platform
  - Architecture
  - Risk Levels
  - Host environment
  - Organizational setup
- Project
  - Stakeholders
  - Departments
  - Time line
  - Team competency
  - Resource,
  - Tools
  - Risk Factors

# Exploratory Testing

## Customer Communication

- What and how is planned forward?
- What have you tested?
- How do we know that you have tested correctly?
- Traceability of test cases to requirements.
- How the test cases can be reused?




- What do you need to know about system?
- What documentation level?
- How much do we need to document?
- Is it reusable?
- Which format the results are required?
- When do you need this information?

# Exploratory Testing

- Major factors affecting in applying ET
  - Explaining to customer
  - Coaching testers
  - Misconceptions
  - Lack of tester's interest
  - Improper planning
  - Managers/customers are afraid of change

# Paper 2

## An Experiment and Survey on the Effectiveness and Efficiency of Exploratory Testing and Traditional Test Case Based Testing

# Research Problem

- **Objects of study** Exploratory and test case based testing.
- **Purpose** Compare and evaluate the testing approaches in terms of found bugs (and bug type) in provided limited time.
- **Quality focus** Defect detection efficiency and effectiveness of testing approaches.
- **Perspective** Tester and research point of view
- **Context** Industry practitioners and academia students. The study is conducted as a Multi-test within object study

# Research Questions

- **RQ-1**:How does using a testing approach (ET or TCT) affect the number of detected defects in limited provided time?
  - ***Null Hypothesis H1.0*** There is no difference in the number of detected defects between ET and TCT.
  - ***Hypothesis H1.1*** More defects are detected with ET than TCT approach
  - ***Hypothesis H1.2*** More defects are detected with TCT than ET approach

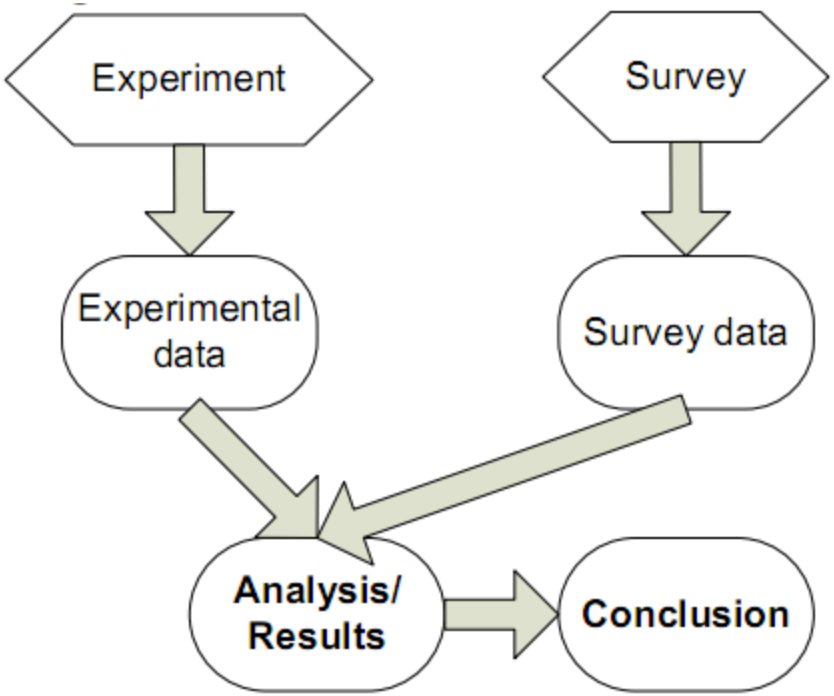
# Research Questions

- **RQ-2**: How does using ET or TCT affect the type of detects identified?
  - ***Null Hypothesis H2.0*** There is no difference in the type of detected defects between using ET or TCT approach

- **RQ-3**: How does using ET or TCT affect the number of false defect reports?
  - ***Null Hypothesis H3.0*** There is no difference in the number of false defect reports between using ET or TCT testing approach

# Research Methods

- Experiment
  - Formal, rigorous and controlled investigation
  - Factors identified and manipulated
  - Provides
    - Execution control
    - Measurement control
    - Ease of replication
- Survey
  - To understand and quantify the perception of approaches



# Experiment Design

## elements

- **Subjects**
  - **Obtain consent**
  - **Sensitive results**

Total 70: 24 from industry and 46 from academia

ET = 35

TCT = 35

| Characteristic | $\bar{x}$ | $\tilde{x}$ | $\sigma$ |
|---|---|---|---|
| SW dev experience (yrs) | 0.583 | 1 | 0.503 |
| Testing experience (yrs) | 0.291 | 0 | 0.464 |

Characteristics of academia subjects

| Characteristic | $\bar{x}$ | $\tilde{x}$ | $\sigma$ |
|---|---|---|---|
| SW dev experience (yrs) | 2.954 | 2 | 2.514 |
| Testing experience (yrs) | 4.045 | 3 | 1.91 |
| TCT based Testing experience (yrs) | 3.5 | 3 | 2.738 |

Characteristics of industry subjects

# Experiment Design

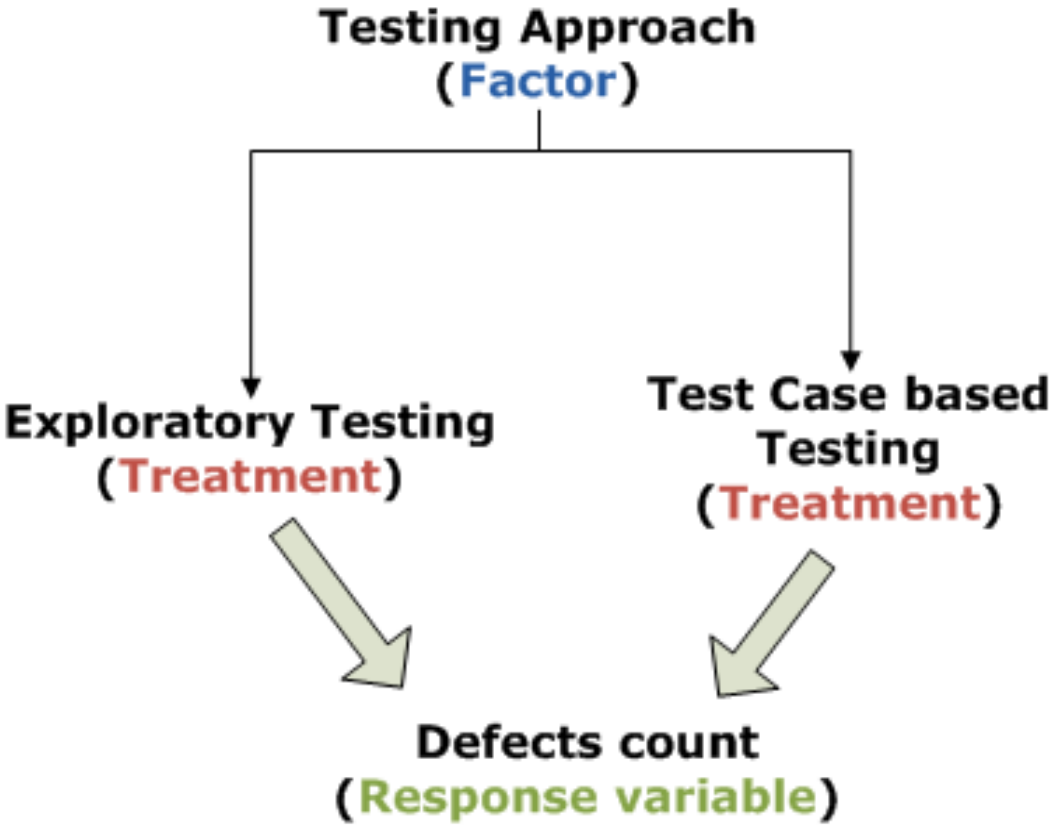
| Phase | ET – group | TCT – group |
|---|---|---|
| Testing Session | Exploratory Testing | Test Case based Testing |

| Phase | Length | Activities |
|---|---|---|
| Session setup | 15 mins | • Instructions<br>• Fill in pre-survey<br>• Access remote environment |
| Functional testing | 90 mins | • Functional testing<br>• Report the bugs |
| Survey and reports handling | 15 mins | • Fill post-survey<br>• Hand over defect reports/logs |

Testing session detail

# Experiment Design

## elements



# Experiment Design

## elements

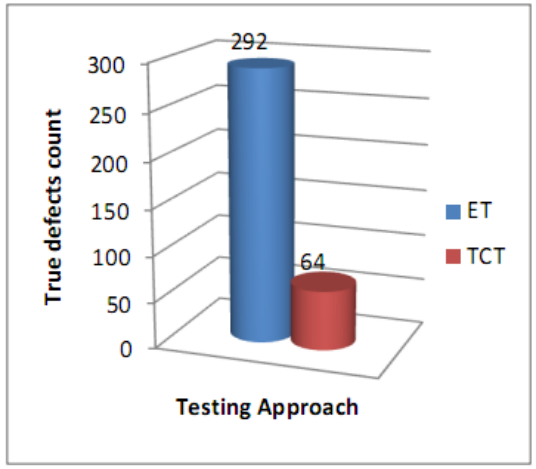
- **Instrumentation**
  - Test object, guidelines, test case design template, defect report, ET charter and survey questionnaires
- **Blocking variables**
  - Actual defects in software, total number & type and difficulty of detecting them
- **Parameters**
  - (Software under test, subject properties, tools, test execution time and test environment)
- **Internal replication**
  - 70 internal replications

# Results

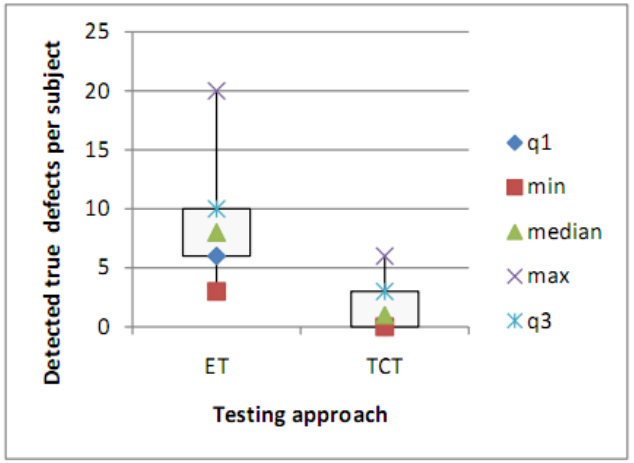
- **Experimental data**
  - True defects
    - Detection difficulty
    - Technical type
    - Defect severity
  - False defects
- **Survey data**
  - Perceived coverage
  - Perceived quality
  - Perception of testing approaches
  - Challenges of testing approach

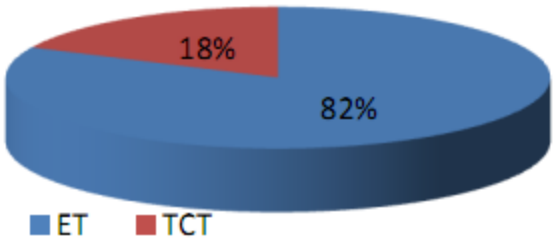
# Experiment Results

- True defects count




| Testing approach | Found defects per subject | | |
|---|---|---|---|
| | $\bar{x}$ | $\sigma$ | $SE_{\bar{x}}$ |
| ET | 8.342 | 4.214 | 0.712 |
| TCT | 1.828 | 1.822 | 0.308 |



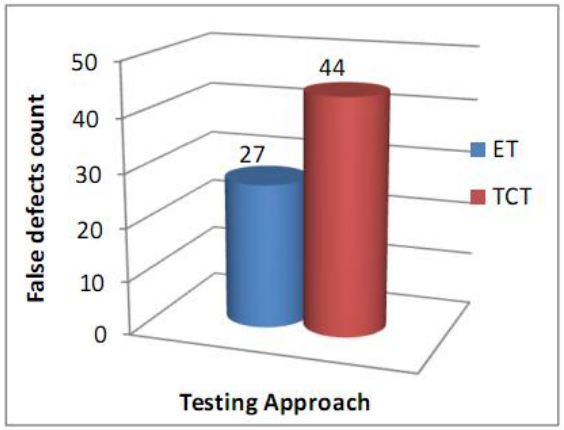




True defect count %

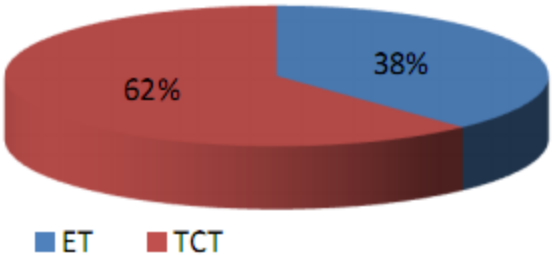
**two-tailed *t*-test**
**statistical significance value *p*= 0.000**

# Experiment Results

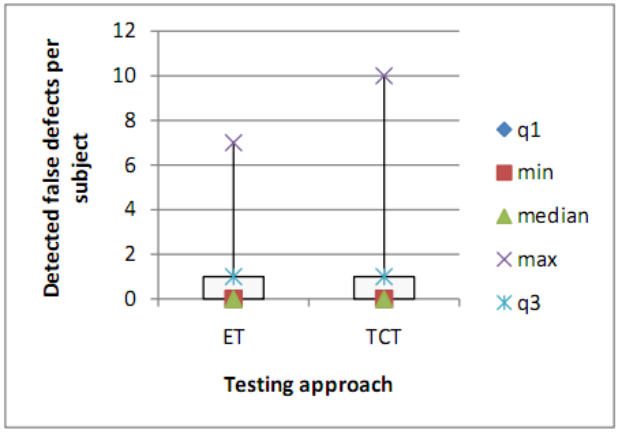
- False defects count




| Testing approach | False defects per subject | |
|---|---|---|
| | $\bar{x}$ | $\sigma$ |
| ET | 0.771 | 1.628 |
| TCT | 1.257 | 2.477 |




False defects count %

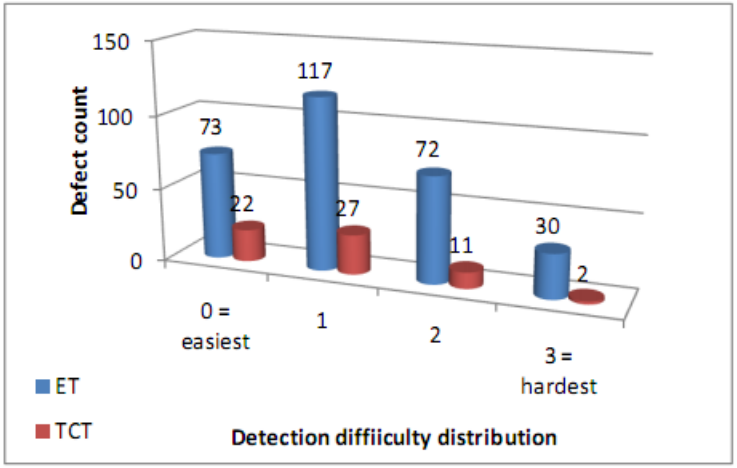



**two-tailed Mann-Whitney statistical significance value $p = 0.584$**

# Experiment Results

- Detection difficulty

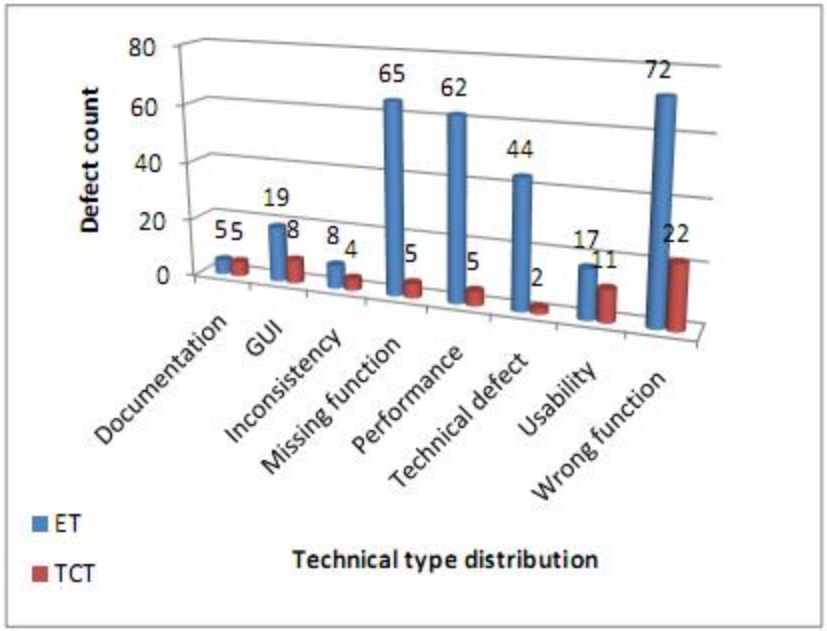
| Mode | ET | TCT | ET/TCT | Total |
|---|---|---|---|---|
| 0 = easiest | 73 | 22 | 331 % | 95 |
| 1 | 117 | 27 | 433 % | 144 |
| 2 | 72 | 11 | 654 % | 83 |
| 3 = hardest | 30 | 2 | 1500 % | 32 |
| Total | 292 | 62 | 470 % | 354 |



# Experiment Results

- Technical type

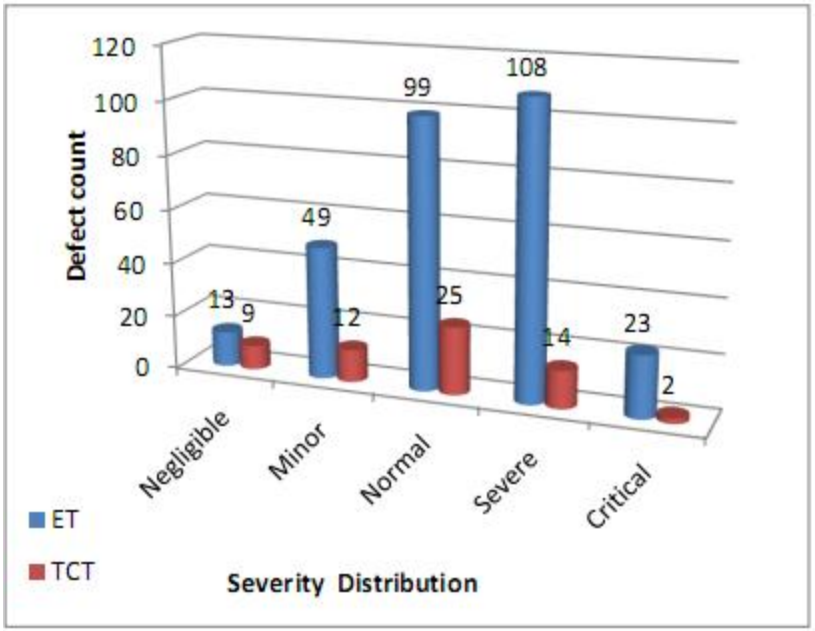
| Technical Type | ET | TCT | ET/TCT | Total |
|---|---|---|---|---|
| Documentation | 5 | 5 | 100 % | 10 |
| GUI | 19 | 8 | 238 % | 27 |
| Inconsistency | 8 | 4 | 200 % | 12 |
| Missing function | 65 | 5 | 1300 % | 70 |
| Performance | 62 | 5 | 1240 % | 67 |
| Technical defect | 44 | 2 | 2200 % | 46 |
| Usability | 17 | 11 | 155 % | 28 |
| Wrong function | 72 | 22 | 327 % | 94 |
| **Total** | 292 | 62 | 471 % | 354 |



# Experiment Results

- Defect severity

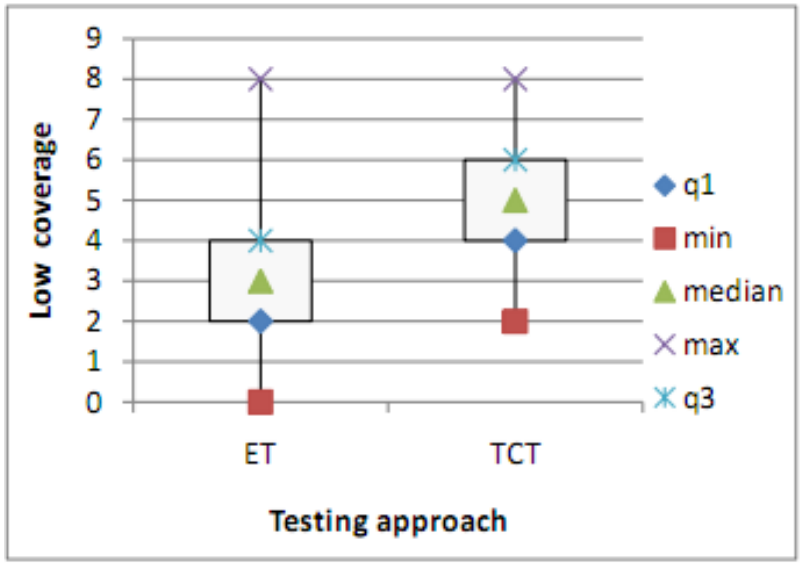
| Severity | ET | TCT | ET/TCT | Total |
|---|---|---|---|---|
| Negligible | 13 | 9 | 144 % | 22 |
| Minor | 49 | 12 | 408 % | 61 |
| Normal | 99 | 25 | 396 % | 124 |
| Severe | 108 | 14 | 771 % | 37 |
| Critical | 23 | 2 | 1150 % | 110 |
| **Total** | 292 | 62 | 471 % | 354 |



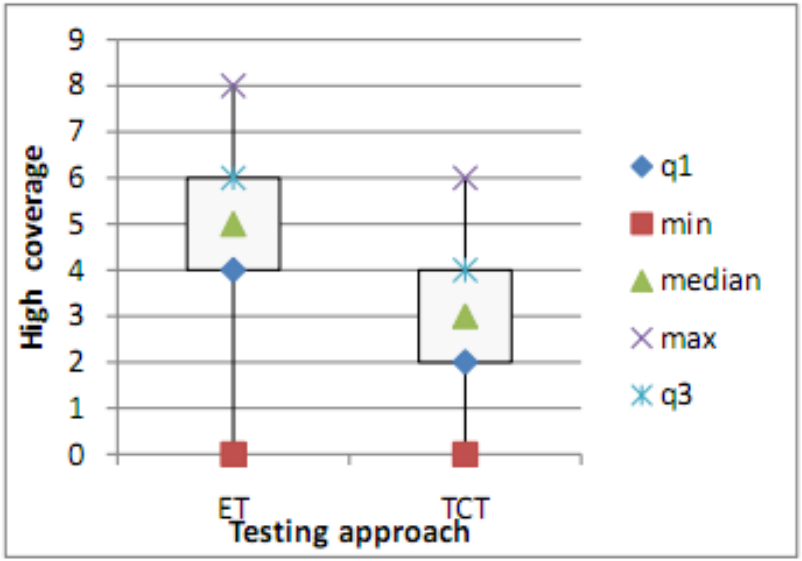
# Survey Results

- Perceived coverage

Not covered at all, Covered superficially




Basic function well covered, covered thoroughly

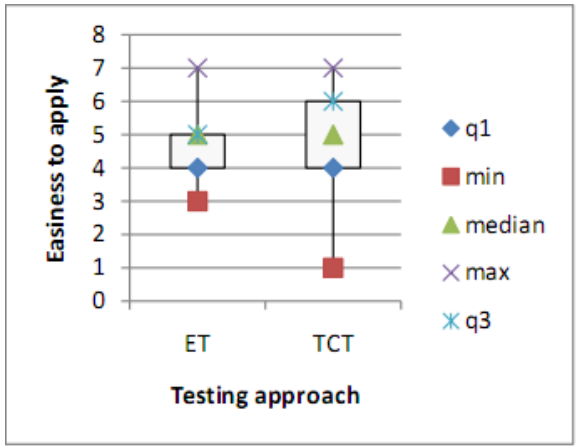



*t*-test *two tailed p = 0.000*

# Survey Results

## Easiness to apply testing approach

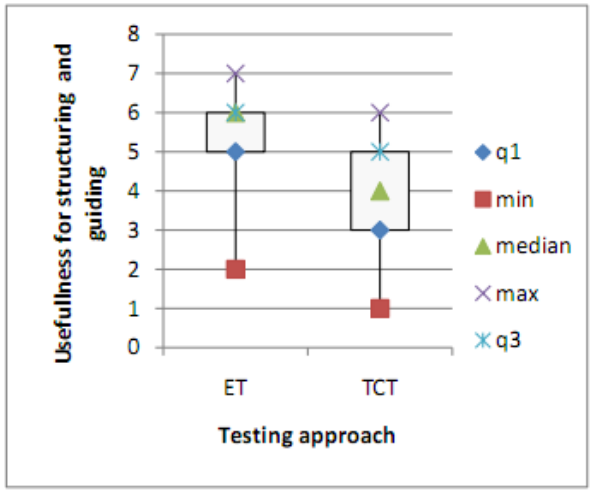
| | Difficult | | | Neutral | | | Very easy |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| ET | 0 | 0 | 5 | 5 | 17 | 7 | 1 |
| TCT | 1 | 0 | 1 | 8 | 11 | 8 | 6 |
| Total | 1 | 0 | 6 | 13 | 28 | 15 | 7 |




*Mann-Whitney test, p = 0.221*

## Usefulness for structuring and guiding

| | Hinder | | | Neutral | | | Very useful |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| ET | 0 | 1 | 0 | 7 | 9 | 12 | 6 |
| TCT | 1 | 2 | 8 | 10 | 9 | 5 | 0 |
| Total | 1 | 3 | 8 | 17 | 18 | 17 | 6 |

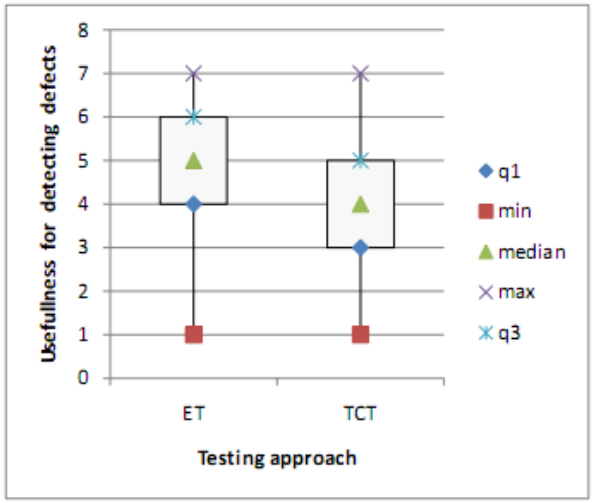



*Mann-Whitney test, p = 0.000*

# Survey Results

## Usefulness for detecting defects

| | Hinder | | | Neutral | | | Very useful |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| ET | 1 | 1 | 4 | 10 | 9 | 8 | 2 |
| TCT | 1 | 3 | 6 | 9 | 9 | 5 | 2 |
| Total | 2 | 4 | 10 | 19 | 18 | 13 | 4 |




*Mann-Whitney test, p = 0.307*

# Contribution

- Systematic Literature Review
- State of the practice knowledge
- Empirical evaluation of defect detection efficiency and effectiveness in exploratory testing

# Thank you!

# Questions?