

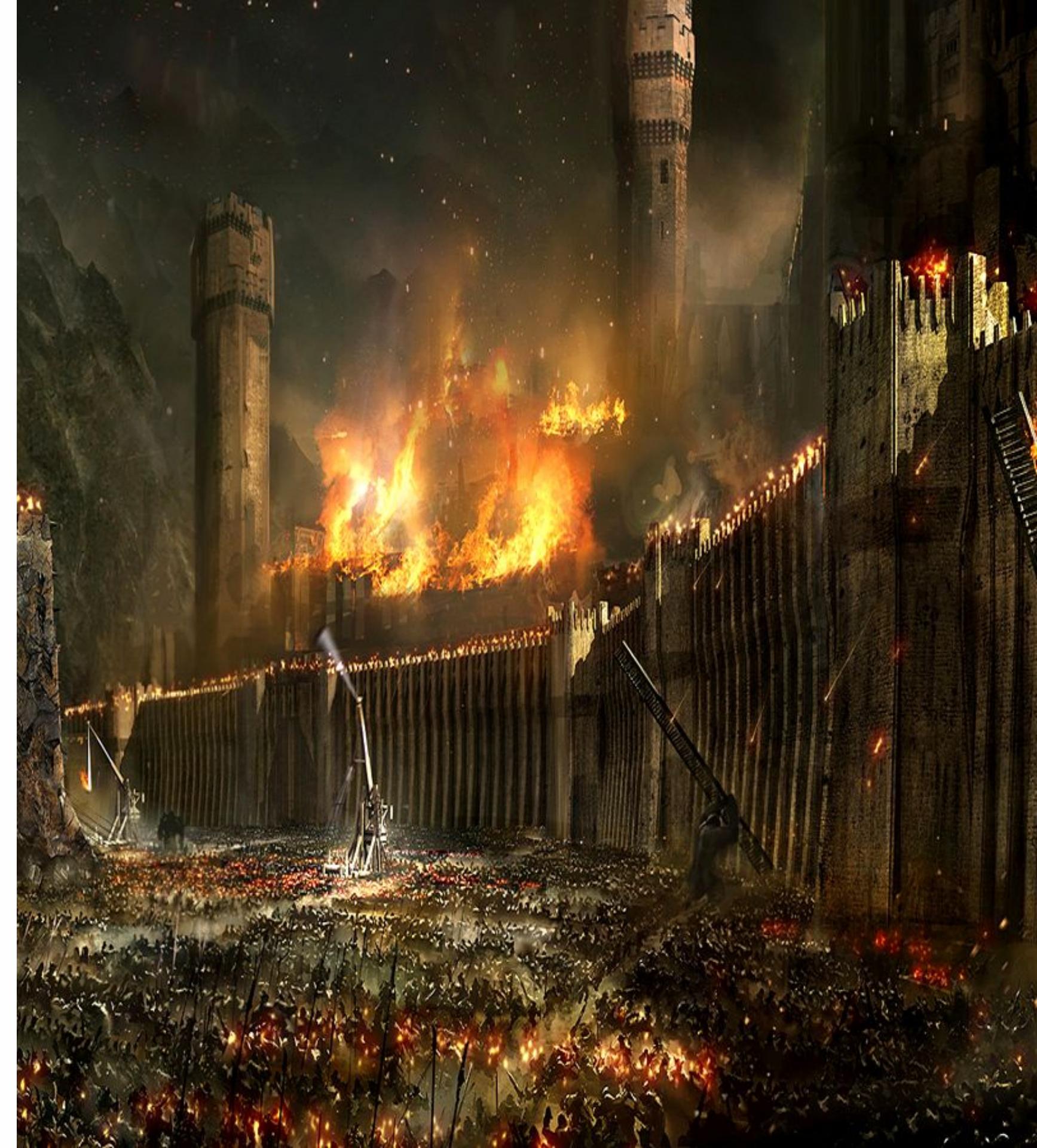
# **AN EMPIRICAL INVESTIGATION ON THE READABILITY OF MANUAL AND GENERATED TEST CASES**

**G. GRANO, S. SCALABRINO, H. C. GALL, R. OLIVETO**

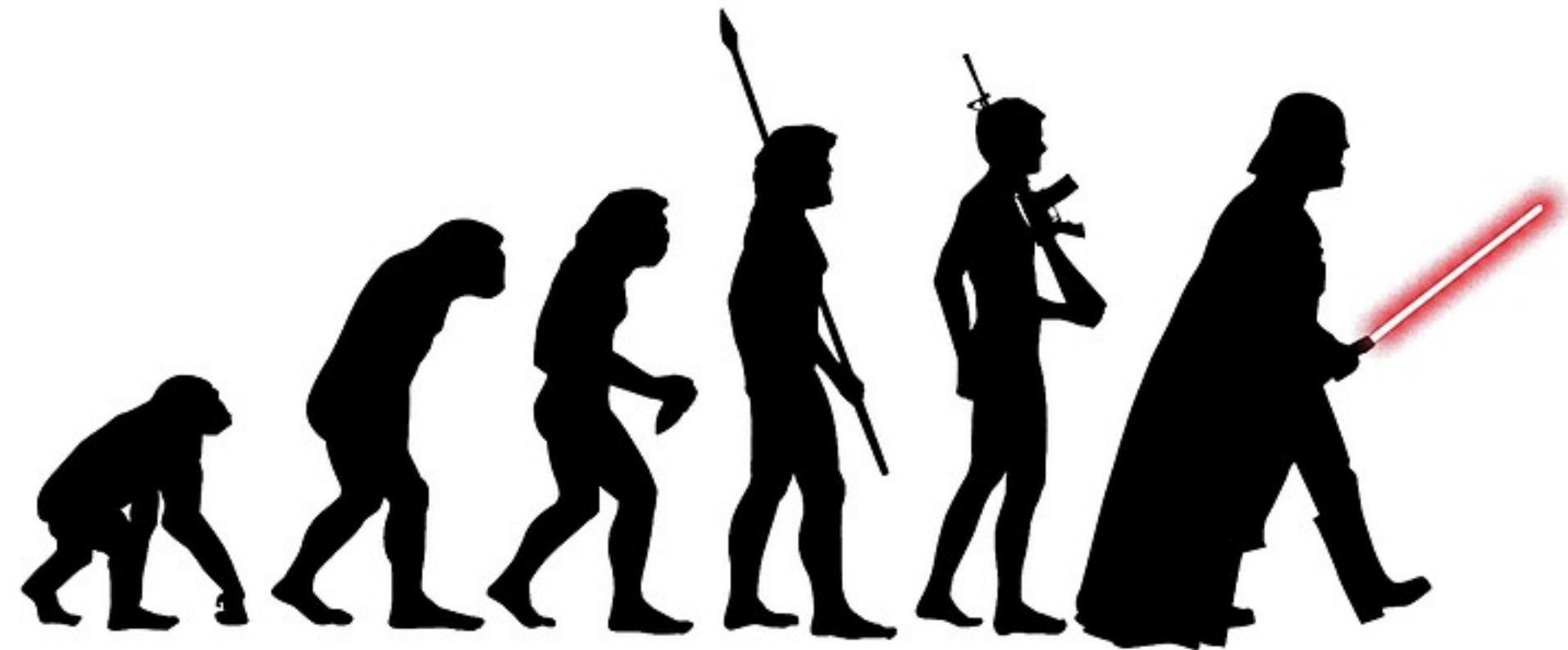
**ICPC ERA 2018, 27-28 MAY, GOTHENBURG (SWEDEN)**

# **UNIT TEST**

## **FIRST DEFENSIVE LINE AGAINST BUG INTRODUCTION**



# TESTS EVOLVE WITH CODE



# TIME SPENT WITH TESTS<sup>1</sup>

- > 416 SOFTWARE ENGINEERS FOR OVER 13 YEARS OF RECORDED WORKING TIME IN THE IDE
- > DEVELOPERS SPEND 25% OF THEIR TIME ENGINEERING TESTS IN THEIR IDE
- > THEY LOOK AT TESTS WHEN ERRORS OCCUR

The screenshot shows the TestRoots website with a navigation bar including Home, WatchDog (which is highlighted in blue), Report, Team, Vacancies, Publications, and an envelope icon. The main content area features a large "WatchDog" logo with a tree graphic. Below the logo, a section titled "DO YOU PROGRAM JAVA WITH ECLIPSE OR INTELLIJ?" contains a snippet of Java code. A call-to-action button at the bottom says "JOIN OVER 4,000 WATCHDOG USERS, WIN AMAZING PRIZES WITH ZERO EXTRA WORK, AND HELP SCIENCE ALONG THE WAY. DOWNLOAD". A note below the button states: "WatchDog is a plugin for Eclipse and IntelliJ IDEs that monitors how you test and debug, gives you helpful statistics on it and it respects your privacy (and your company's policy)."

<sup>1</sup>BELLER ET.AL - WHEN, HOW, AND WHY DEVELOPERS (DO NOT) TEST IN THEIR IDES

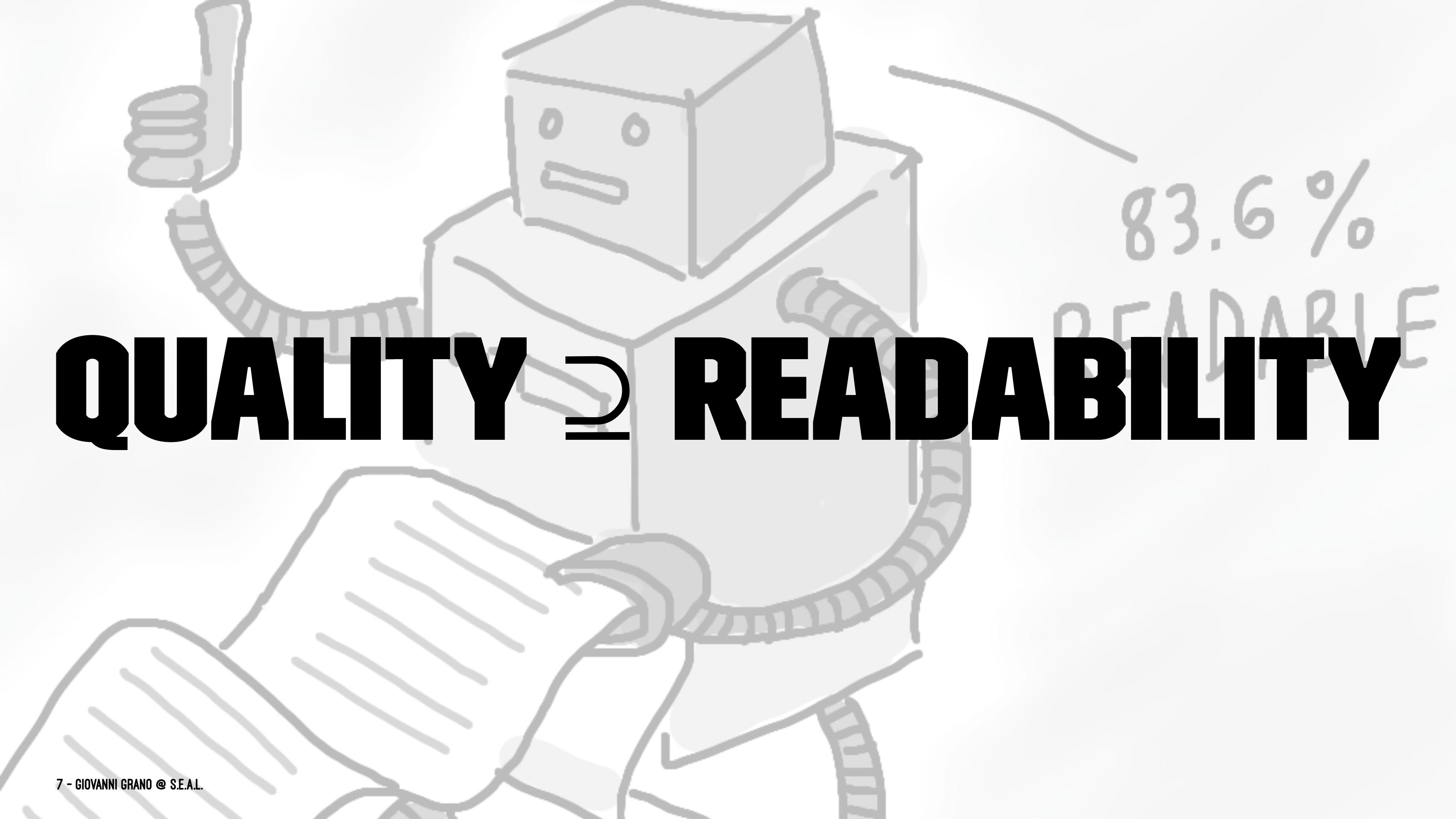
# **SAME LEVEL OF QUALITY**

## **SAME EFFORT IN WRITING GOOD QUALITY TESTS AND PRODUCTION CODE**

- › CHANGES ARE EASIER
- › ERROR COMPREHENSION

A black and white photograph of a person from behind, wearing a dark hoodie. The word "FIRST" is printed in large, light-colored letters on the back of the hoodie.

# **ARE TESTS FIRST-CLASS CITIZENS?**



# **QUALITY → READABILITY**

# STUDY DESIGN

- > 3 APACHE OPEN-SOURCE PROJECTS
- > STATE-OF-THE-ART READABILITY MODEL <sup>2</sup>
- > PAIRS OF <TEST-PRODUCTION CODE>
- > WILCOXON TEST

<sup>2</sup> SCALABRINO ET.AL. - IMPROVING CODE READABILITY MODELS WITH TEXTUAL FEATURES - ICPC 2016

# FEATURES

- › STRUCTURAL FEATURES (E.G. LINE LENGTH)
- › VISUAL FEATURES (E.G. INDENTATION)
- › TEXT-BASED CODE READABILITY FEATURES (E.G. COMMENTS AND IDENTIFIERS CONSISTENCY)

# RQ1

## ARE TEST CASES AND CLASSES EQUALLY READABLE?

CUT	TESTS	P-VALUE	MAGNITUDE
.74	.50	.009	LARGE
.86	.64	.000	LARGE
.82	.61	.000	LARGE
.83	.61	.000	LARGE

**TEST CASES ARE LESS  
READABLE THAN THE CODE  
THEY EXERCISE**

**WELL... WE HAVE TOOLS!**



# RQ2

## ARE GENERATED AND MANUAL TESTS EQUALLY READABLE?

MANUAL	AUTOMATED	P-VALUE	MAGNITUDE
.50	.48	.791	
.64	.53	.000	MEDIUM
.61	.52	.000	SMALL
.61	.52	.000	SMALL

**MANUALLY WRITTEN TESTS  
ARE MORE READABLE!**

# FUTURE WORK

## WHY READABILITY OF TEST IS LOWER?

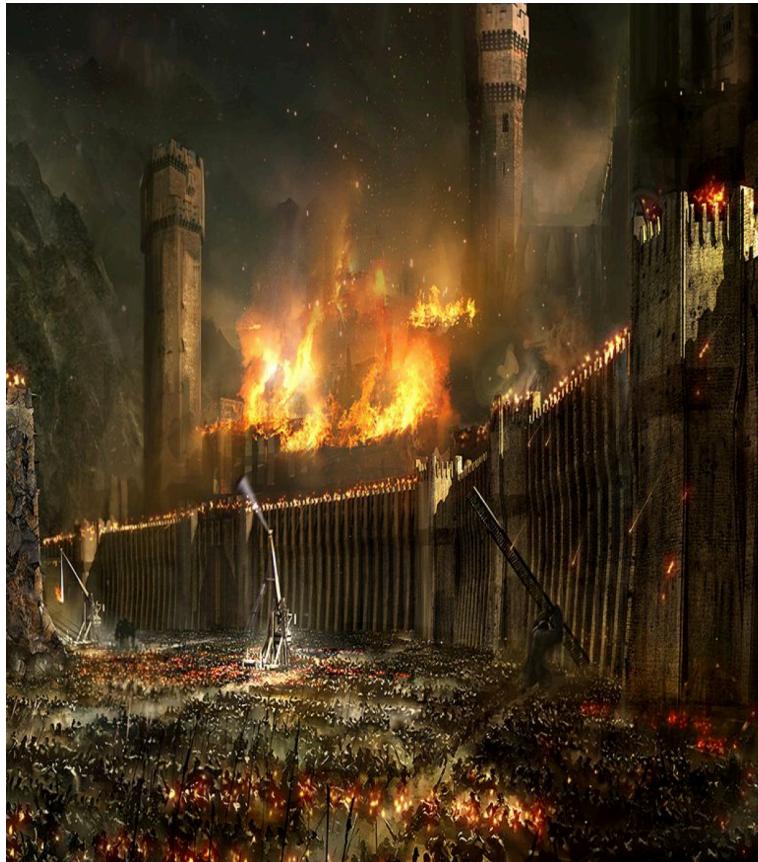
- › GOOD PROXY FOR QUALITY?
- › INCREASE AWARENESS
- › SUPPORT TEST QUALITY

## CAN WE IMPROVE THE TOOLS?



# UNIT TEST FIRST DEFENSIVE LINE AGAINST BUG INTRODUCTION

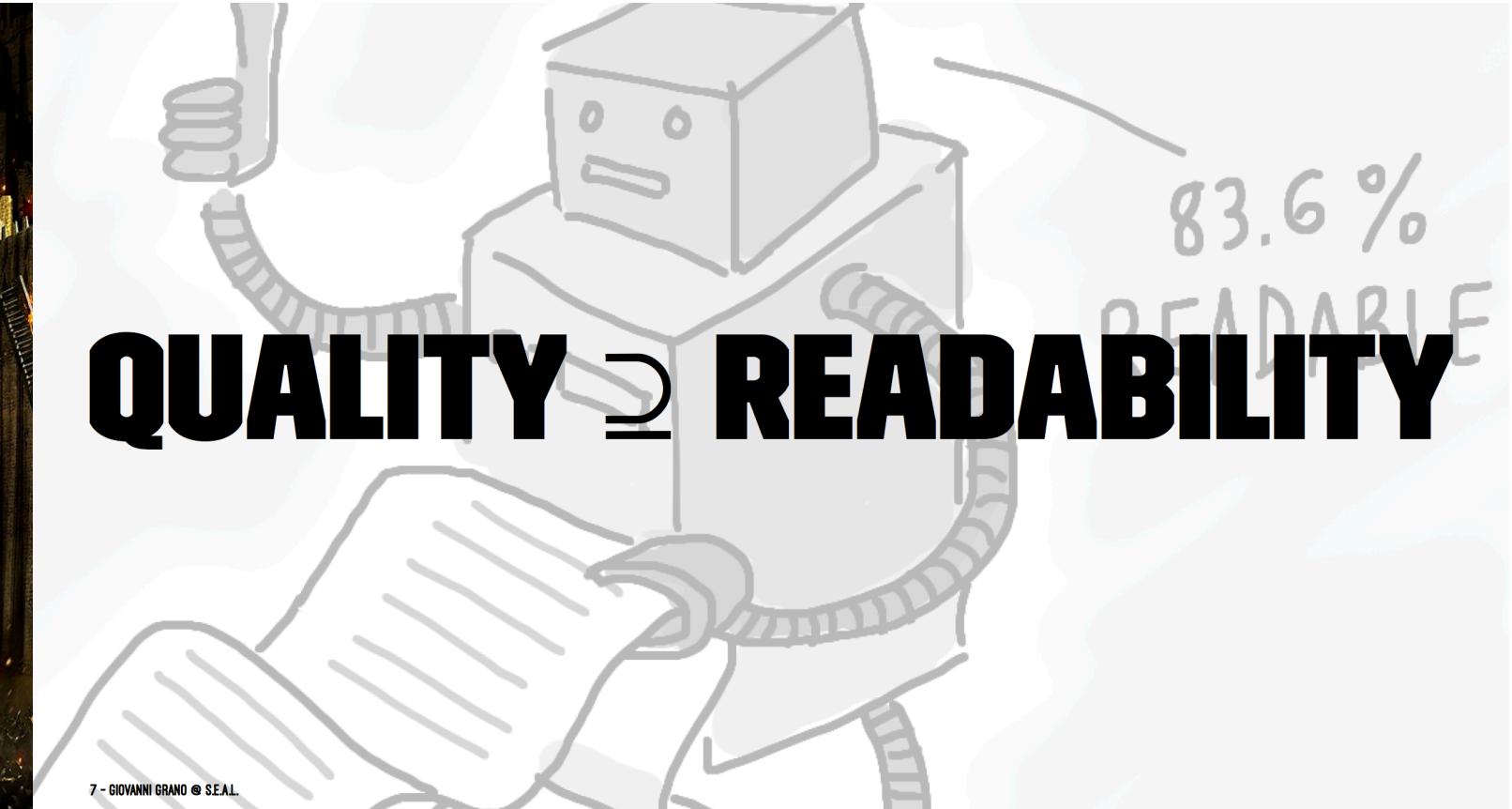
2 - GIOVANNI GRANO @ S.E.A.L.



CUT	TESTS	P-VALUE	MAGNITUDE
.74	.50	.009	LARGE
.86	.64	.000	LARGE
.82	.61	.000	LARGE
.83	.61	.000	LARGE

10 - GIOVANNI GRANO @ S.E.A.L.

**TEST CASES ARE LESS  
READABLE THAN THE CODE  
THEY EXERCISE**



**WELL... WE HAVE TOOLS!**

8 - GIOVANNI GRANO @ S.E.A.L.