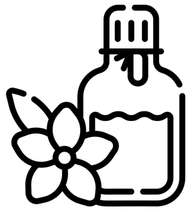
# Pizza vs Pinsa:
# On the Perception and Measurability of Unit Test Code Quality

Giovanni Grano, Christian De Iaco, Fabio Palomba, Harald C. Gall

36th IEEE International Conference on Software Maintenance and Evolution

# first and foremost: what is a pinsa?

# motivation



essential asset to foster software quality

tests of good quality

do existing metrics match developers’ perception of test quality?

# right metrics

oil

water

flour

# goals and research questions

2 research questions

- what are the features that influence unit test quality?
- do existing metrics match developers’ perception?

mixed-method research approach

# RQ1: the practitioner’s perspective

interviews

5 experts

~ 60 minutes

initial taxonomy

- general discussion
- own definition of test quality
- summarize measurable factors

# RQ1: the practitioner’s perspective

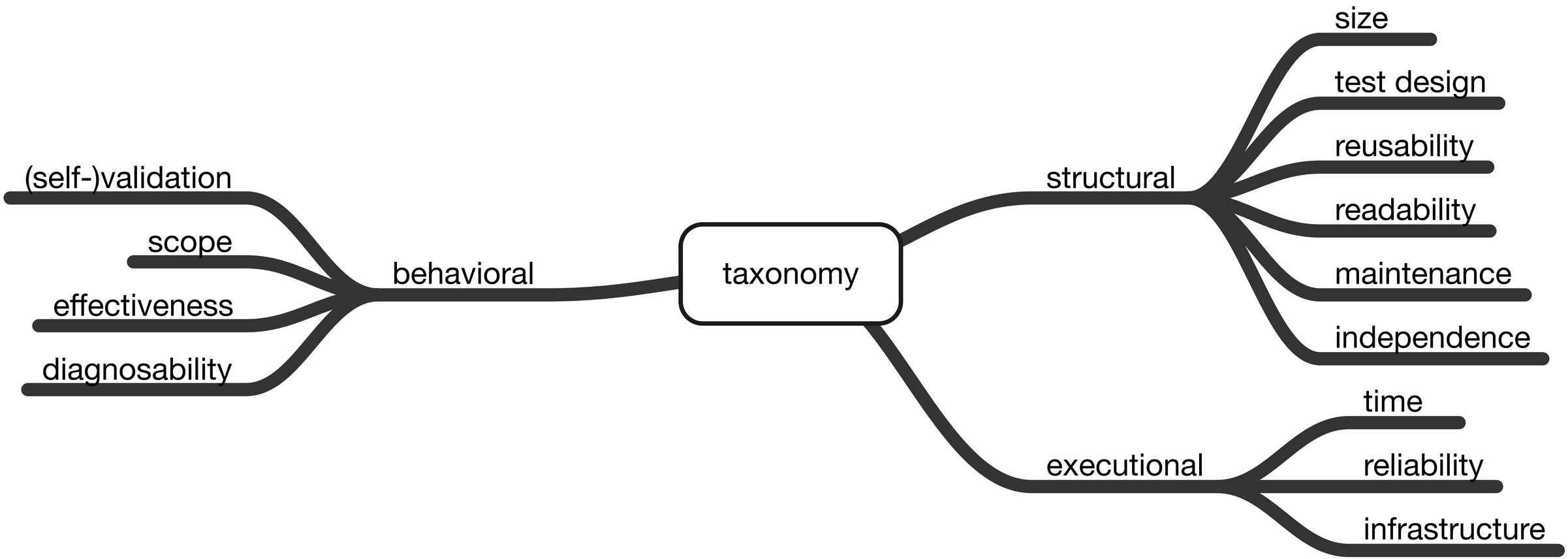
survey

70 developer

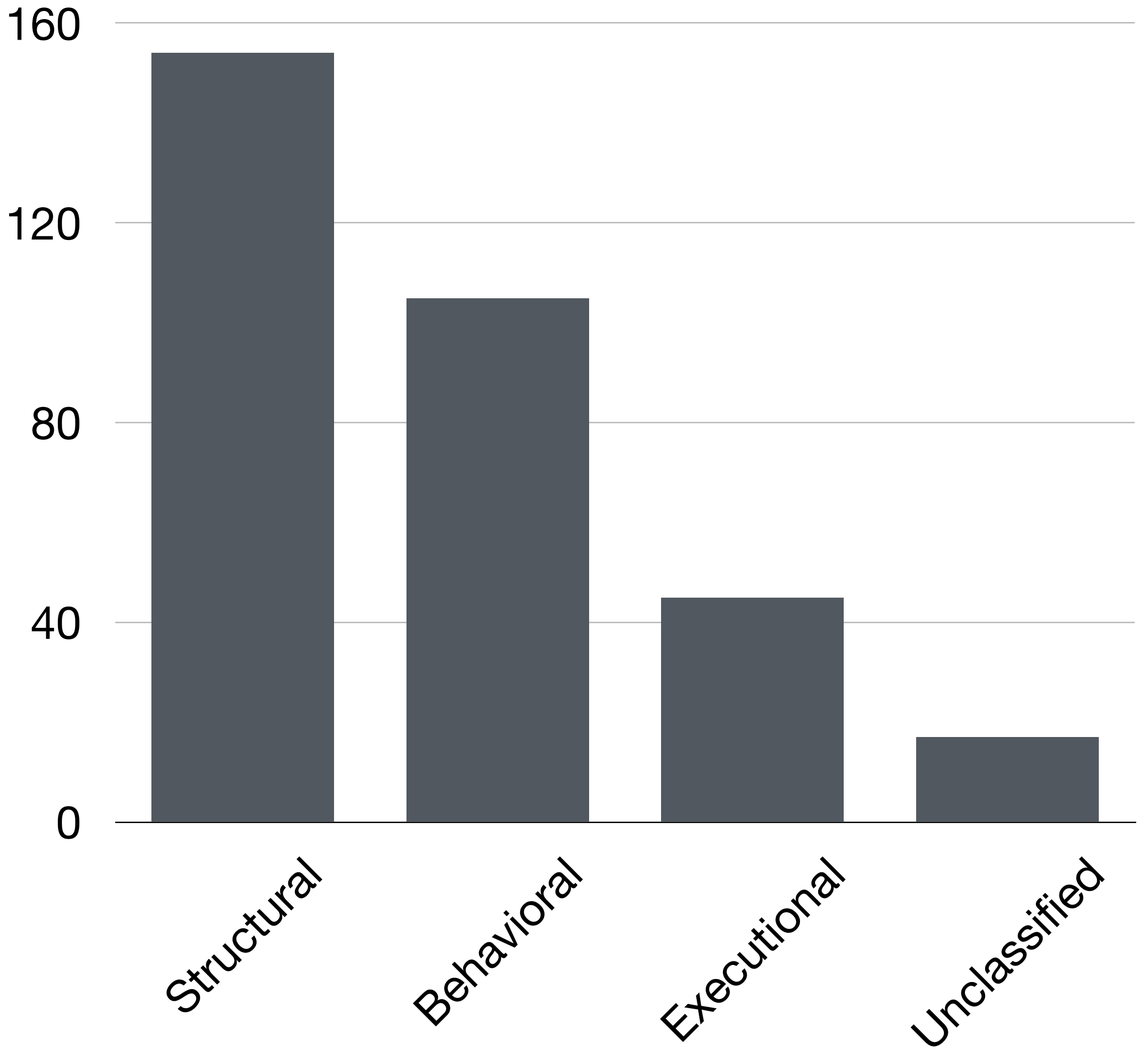
march 2020

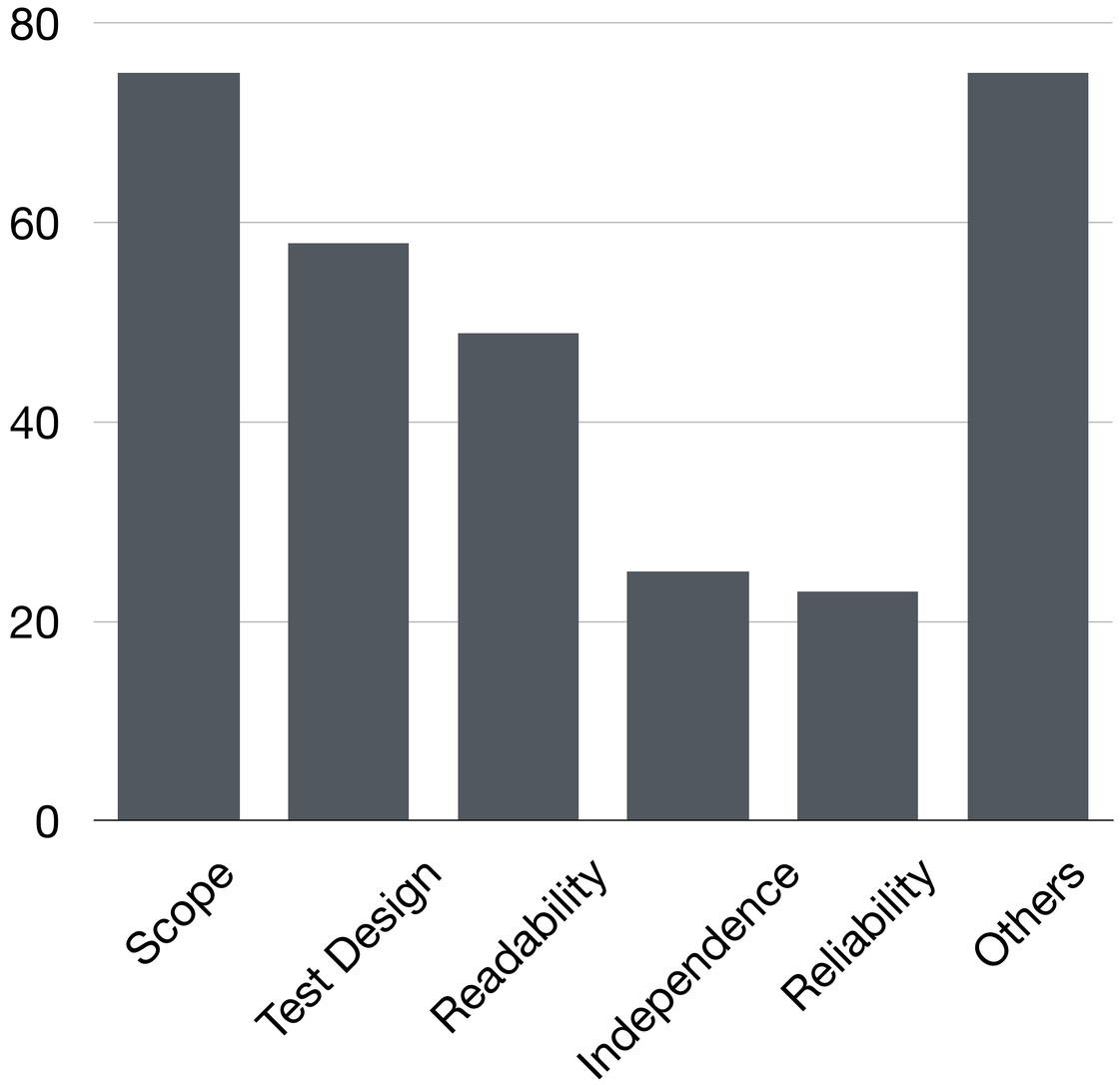
- confirm initial taxonomy
- additional factors
- rate quality of test snippets

# RQ1: results


taxonomy
behavioral
(self-)validation
scope
effectiveness
diagnosability
structural
size
test design
reusability
readability
maintenance
independence
executional
time
reliability
infrastructure

# RQ1: results


160
120
80
40
0
Structural
Behavioral
Executional
Unclassified


80
60
40
20
0
Scope
Test Design
Readability
Independence
Reliability
Others

# RQ1: findings

multi-faceted concept


non-functional aspects

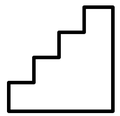
test code design

# RQ2: the research perspective

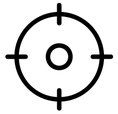
survey

70 developer


10 test code snippets

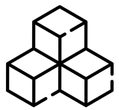


1-5 Likert scale



response variable:
score on the Likert scale

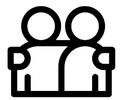

independent variables:
11 metrics



proportional odds model

# RQ2: results

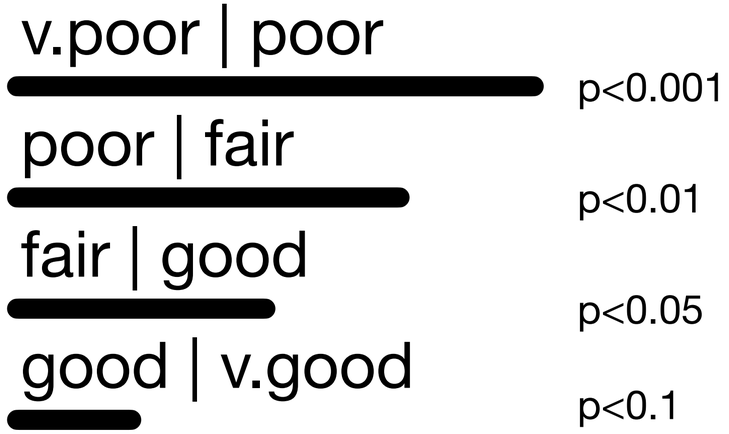
mutation score



coupling between objects

assertion roulette

number of static invocations



# RQ2: findings

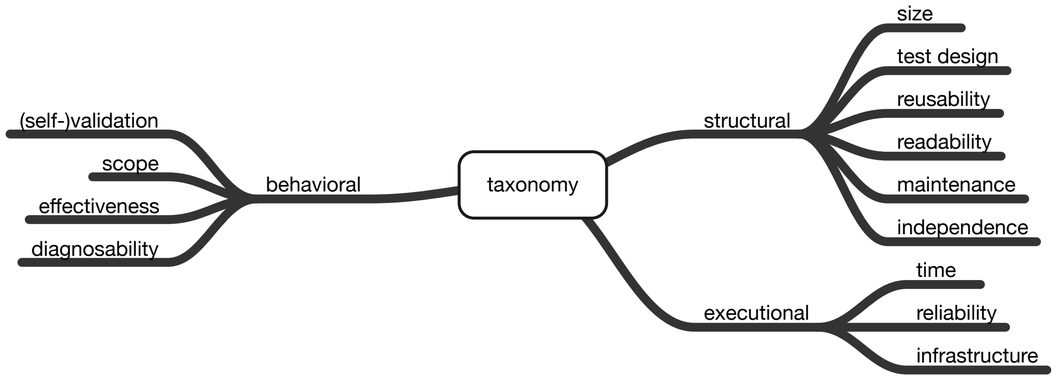
metrics can discern low-quality from fair ones


poor ability of metrics representing size, complexity, and readability

metrics fails at providing a comprehensive model of perceived test quality

## first and foremost: what is a pinsa?

## RQ1: results




## RQ2: findings

- metrics can discern low-quality from fair ones
- poor ability of metrics representing size, complexity, and readability
- metrics fails at providing a comprehensive model of perceived test quality

@giograno90