

# Investigating the Impact of Development Task on External Quality in Test-Driven Development: An Industry Experiment

Ayşe Tosun<sup>1</sup>, Oscar Dieste<sup>2</sup>, Sira Vegas<sup>2</sup>, Dietmar Pfahl<sup>3</sup>, Kerli Rungi<sup>4</sup>, and Natalia Juristo<sup>2,5</sup>

<sup>1</sup>Faculty Istanbul Technical University, Turkey [tosunay@itu.edu.tr](mailto:tosunay@itu.edu.tr)

<sup>2</sup>Universidad Politécnica de Madrid, Spain [{odieste,svegas,njuristo}@fi.upm.es](mailto:{odieste,svegas,njuristo}@fi.upm.es)

<sup>3</sup>University of Tartu, Estonia [dietmar.pfahl@ut.ee](mailto:dietmar.pfahl@ut.ee)

<sup>4</sup>Testlio OY, Estonia [kerli.rungi@gmail.com](mailto:kerli.rungi@gmail.com)

<sup>5</sup>University of Oulu, Finland [natalia.juristo@oulu.fi](mailto:natalia.juristo@oulu.fi)

**Palabras clave:** Test-driven development, industry experiment, experimental task, incremental test-last development, external quality.

**Lugar de publicación:** *IEEE Trans. Softw. Eng.*, 47(11): 2438-2456 (2021).

**Índice de impacto:** JCR 6,226 - Cuartil Q1 - Posición: 7/108 - Área: Computer Science/Software Engineering.

**DOI:** 10.1109/TSE.2019.2949811

**Abstract.** Reviews on test-driven development (TDD) studies suggest that the conflicting results reported in the literature are due to unobserved factors, such as the tasks used in the experiments, and highlight that there are very few industry experiments conducted with professionals. The goal of this study is to investigate the impact of a new factor, the chosen *task*, and the *development approach* on external quality in an industrial experimental setting with 17 professionals. The participants are junior to senior developers in programming with Java, beginner to novice in unit testing, JUnit, and they have no prior experience in TDD. The experimental design is a  $2 \times 2$  cross-over, i.e., we use two *tasks* for each of the two approaches, namely TDD and incremental test-last development (ITLD). Our results reveal that both *development approach* and *task* are significant factors with regards to the external quality achieved by the participants. More specifically, the participants produce higher quality code during ITLD in which splitting user stories into subtasks, coding, and testing activities are followed, compared to TDD. The results also indicate that the participants produce higher quality code during the implementation of Bowling Score Keeper, compared to that of Mars Rover API, although they perceived both *tasks* as of similar complexity. An interaction between the *development approach* and *task* could not be observed in this experiment. We conclude that variables that have not been explored so often, such as the extent to which the *task* is specified in terms of smaller subtasks, and developers' unit testing experience might be critical factors in TDD experiments. The real-world appliance of TDD and its implications on external quality still remain to be challenging unless these uncontrolled and unconsidered factors are further investigated by researchers in both academic and industrial settings.