Analyzing The Effects of Test Driven Development In GitHub

Neil C. Borle
Department of Computing Science,
University of Alberta
Edmonton, Alberta, Canada
nborle@ualberta.ca

Meysam Feghhi
Department of Computing Science,
University of Alberta
Edmonton, Alberta, Canada
feghhi@ualberta.ca

Eleni Stroulia
Department of Computing Science,
University of Alberta
Edmonton, Alberta, Canada
stroulia@ualberta.ca

Russell Greiner
Department of Computing Science,
University of Alberta
Edmonton, Alberta, Canada
rgreiner@ualberta.ca

Abram Hindle
Department of Computing Science,
University of Alberta
Edmonton, Alberta, Canada
hindle1@ualberta.ca

ABSTRACT

Testing is an integral part of the software development lifecycle, approached with varying degrees of rigor by different process models. Agile process models recommend Test Driven Development (TDD) as a key practice for reducing costs and improving code quality. The objective of this work is to perform a cost-benefit analysis of this practice. Previous work by Fucci et al. [2, 3] engaged in laboratory studies of developers actively engaged in test-driven development practices. Fucci et al. found little difference between test-first behaviour of TDD and test-later behaviour. To that end, we opted to conduct a study about TDD behaviours in the "wild" rather than in the laboratory. Thus we have conducted a comparative analysis of GitHub repositories that adopts TDD to a lesser or greater extent, in order to determine how TDD affects software development productivity and software quality. We classified GitHub repositories archived in 2015 in terms of how rigorously they practiced TDD, thus creating a TDD spectrum. We then matched and compared various subsets of these repositories on this TDD spectrum with control sets of equal size. The control sets were samples from all GitHub repositories that matched certain characteristics, and that contained at least one test file. We compared how the TDD sets differed from the control sets on the following characteristics: number of test files, average commit velocity, number of bug-referencing commits, number of issues recorded, usage of continuous integration, number of pull requests, and distribution of commits per author. We found that Java TDD projects were relatively rare. In addition, there were very few significant differences in any of the metrics we used to compare TDD-like and non-TDD projects; therefore, our results do not reveal any observable benefits from using TDD.

KEYWORDS

Test Driven Development, Human Factors in Software Development, GitHub Repositories, Continuous Integration

ICSE '18, May 27-June 3, 2018, Gothenburg, Sweden

ACM ISBN 978-1-4503-5638-1/18/05.
https://doi.org/10.1145/3180155.3182535

ACM Reference Format:
Neil C. Borle, Meysam Feghhi, Eleni Stroulia, Russell Greiner, and Abram Hindle. 2018. Analyzing The Effects of Test Driven Development In GitHub. In *ICSE '18: ICSE '18: 40th International Conference on Software Engineering , May 27-June 3, 2018, Gothenburg, Sweden.* ACM, New York, NY, USA, Article 1, 1 page. https://doi.org/10.1145/3180155.3182535

Test-driven development within the Java GitHub ecosystem is quite rare.

Projects that engaged in test-driven development did not show any significant differences in terms of quality over those projects that simply engaged in testing.

This work [1] was published by Springer in the journal of Empirical Software Engineering on November 25, 2017. The paper is available from the publisher's website at:
http://rdcu.be/zqhX
The paper [1] is available as a preprint at:
http://softwareprocess.es/pubs/borle2017EMSE-TDD.pdf
A blog post describing the paper is available at:
http://softwareprocess.es/blog/blog/2017/11/26/analyzing-the-effects-of-test-driven-development-in-github/

REFERENCES

[1] Neil C. Borle, Meysam Feghhi, Eleni Stroulia, Russell Greiner, and Abram Hindle. 2017. Analyzing the effects of test driven development in GitHub. *Empirical Software Engineering* (25 Nov 2017). https://doi.org/10.1007/s10664-017-9576-3 Preprint available at http://softwareprocess.es/pubs/borle2017EMSE-TDD.pdf Publisher URL: http://rdcu.be/zqhX Blog Post: http://softwareprocess.es/blog/blog/2017/11/26/analyzing-the-effects-of-test-driven-development-in-github/.

[2] Davide Fucci, Hakan Erdogmus, Burak Turhan, Markku Oivo, and Natalia Juristo. 2016. A Dissection of Test-Driven Development: Does It Really Matter to Test-First or to Test-Last? *IEEE Transactions on Software Engineering* (2016).

[3] Davide Fucci, Giuseppe Scanniello, Simone Romano, Martin Shepperd, Boyce Sigweni, Fernando Uyaguari, Burak Turhan, Natalia Juristo, and Markku Oivo. 2016. An external replication on the effects of test-driven development using a multi-site blind analysis approach. In *Proceedings of the 10th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement.* ACM, 3.