Information and Software Technology 000 (2017) 1–27

Contents lists available at ScienceDirect

Information and Software Technology

journal homepage: www.elsevier.com/locate/infsof

Software test maturity assessment and test process improvement: A multivocal literature review

Vahid Garousi[a,b,*], Michael Felderer[c], Tuna Hacaloğlu[d,e]

[a] Software Engineering Research Group, Department of Computer Engineering, Hacettepe University, Ankara, Turkey
[b] Maral Software Engineering Consulting, Corporation, Calgary, Canada
[c] Quality Engineering Research Group, Institute of Computer Science, University of Innsbruck, Innsbruck, Austria
[d] Informatics Institute, Middle East Technical University (METU), Ankara, Turkey
[e] Department of Information Systems Engineering, Atilim University, Ankara, Turkey

ARTICLE INFO

Article history:
Received 7 August 2016
Revised 23 November 2016
Accepted 4 January 2017
Available online xxx

Keywords:
Software testing
Test management
Test process
Test maturity
Test process assessment
Test process improvement
Multivocal literature review
Systematic literature review

ABSTRACT

Context: Software testing practices and processes in many companies are far from being mature and are usually conducted in ad-hoc fashions. Such immature practices lead to various negative outcomes, e.g., ineffectiveness of testing practices in detecting all the defects, and cost and schedule overruns of testing activities. To conduct test maturity assessment (TMA) and test process improvement (TPI) in a systematic manner, various TMA/TPI models and approaches have been proposed.

Objective: It is important to identify the state-of-the-art and the –practice in this area to consolidate the list of all various test maturity models proposed by practitioners and researchers, the drivers of TMA/TPI, the associated challenges and the benefits and results of TMA/TPI. Our article aims to benefit the readers (both practitioners and researchers) by providing the most comprehensive survey of the area, to this date, in assessing and improving the maturity of test processes.

Method: To achieve the above objective, we have performed a Multivocal Literature Review (MLR) study to find out what we know about TMA/TPI. A MLR is a form of a Systematic Literature Review (SLR) which includes the grey literature (e.g., blog posts and white papers) in addition to the published (formal) literature (e.g., journal and conference papers). We searched the academic literature using the Google Scholar and the grey literature using the regular Google search engine.

Results: Our MLR and its results are based on 181 sources, 51 (29%) of which were grey literature and 130 (71%) were formally published sources. By summarizing what we know about TMA/TPI, our review identified 58 different test maturity models and a large number of sources with varying degrees of empirical evidence on this topic. We also conducted qualitative analysis (coding) to synthesize the drivers, challenges and benefits of TMA/TPI from the primary sources.

Conclusion: We show that current maturity models and techniques in TMA/TPI provides reasonable advice for industry and the research community. We suggest directions for follow-up work, e.g., using the findings of this MLR in industry-academia collaborative projects and empirical evaluation of models and techniques in the area of TMA/TPI as reported in this article.

1. Introduction

Software testing is an impotent while a costly phase of the software development life-cycle. A 2013 study by the Cambridge University [1] states that the global cost of detecting and fixing software defects has risen to $312 billion annually and it makes up half of the development time of the average project.

According to various studies, e.g., [2–4], software testing practices and processes in many companies are far from being mature and are usually conducted in ad-hoc fashions. Such immature practices lead to various negative outcomes, e.g., ineffectiveness of testing practices in detecting all the defects, and cost and schedule overruns of testing activities. Also, testing is often conduct not efficiently, e.g., "*The costs of testing of a software project or product are considerable and therefore it is important to identify process improvement propositions for testing*" [5].

To conduct Test Maturity Assessment (TMA) and Test Process Improvement (TPI), together referred to as TMA/TPI, in a systematic manner, various TMA/TPI approaches and frameworks have been

* Corresponding author.
E-mail addresses: vahid.garousi@hacettepe.edu.tr (V. Garousi), michael.felderer@uibk.ac.at (M. Felderer), tuna.hacaloglu@metu.edu.tr (T. Hacaloğlu).

http://dx.doi.org/10.1016/j.infsof.2017.01.001
0950-5849/

proposed. For example, a 2014 book entitled "*Improving the Test Process: Implementing Improvement and Change*" [6] by the International Software Testing Qualifications Board (ISTQB) suggest various approaches in this context.

To identify the state-of-the-art and the –practice in this area of scientific and practical interest and to find out what we know about TMA/TPI, we report in this work a 'multivocal' literature review on both the scientific literature and also practitioners' grey literature and we present its results in this article. A multivocal literature review (MLR) [7–9] is a type of systematic literature reviews (SLR) in which data from various sources are included, e.g., scientific literature and practitioners' grey literature (e.g., blog posts, white papers, and presentation videos). We believe that conducting a MLR in the area of TMA/TPI will be more useful compared to a SLR since there is a large of body of knowledge and experience reported by practitioners in the grey literature (e.g., blog posts and white papers) which a regular SLR study will not review and synthesize (by being limited to only the formal published literature).

MLRs have recently started to appear in software engineering, e.g., in recent ones in the areas of technical debt [7] and test automation [10], respectively. Furthermore, the need for more MLRs in software engineering has recently been pointed out and investigated empirically [11], especially also by pointing to the field of test process improvement, which is of high interest to research and practice. By summarizing what we know about TMA/TPI, our systematic review identified 58 different test maturity models and a large number of sources with varying degrees of empirical evidence on this topic. Our article aims to benefit the readers (both practitioners and researchers) in assessing and improving the maturity of test processes by benefitting the state-of-the-art and the –practice in this area.

While there exist a few review (survey) papers on the topic of TMA/TPI, e.g. [12, 13], none of the existing surveys have considered both the academic literature and the practitioners' grey literature and also in the depth that we have conducted in this study, by identifying 58 test maturity models and also the drivers, challenges and benefits of TMA/TPI.

On another note, we would like to clearly note the scope of this study before continuing with the rest of the study. We are aware that testing is not the only approach for software quality assurance, verification and validation (V&V). Techniques such as formal methods, inspections, static code analysis and peer reviews as other forms of V&V that are complementary to testing. But to keep our MLR study focused, we have only focused on surveying the maturity assessment and process improvement approaches specific to software testing and have excluded those focusing on the other V&V activities, e.g., studies such as [14, 15]. Certainly, we encourage MLR and other types of review studies on those other subareas of software quality assurance.

The remainder of this article is structured as follows. A review of the related work is presented in Section 2. We describe the study goal and research methodology in Section 3. Section 4 presents the searching phase and selection of sources. Section 5 discusses the development of the systematic map and data-extraction plan. Section 6 presents the results of the MLR. Section 7 summarizes the findings and discusses the lessons learned. Finally, in Section 8, we draw conclusions, and suggest areas for further research.

2. Background and related work

In this section, we first provide a brief overview of the technical domain of this MLR (software test maturity assessment and test process improvement). We then briefly provide a background on multivocal literature reviews (MLRs) since it is a relatively new terminology in SE. We finish the section by reviewing the related work, i.e., other secondary studies in the scope of TMA/TPI and how our work differs from them and contributes new knowledge to the literature.

2.1. Brief overview of software test maturity assessment and test process improvement

Software testing practices and processes in many companies are far from mature and are still usually conducted in ad-hoc fashions [2–4]. Thus, many team and companies are interested to assess and improve the maturity of their software testing practices and processes.

A recent 2016 SLR [12] on this topic identified 18 TPI approaches showing the fast progress of this important field in software testing. According to many sources (e.g., [12]), TMMi [16, 17] and TPI [18] (and its newer version TPI-Next [19]) are the most popular and widely-used models and approaches in this area. We provide a brief overview of TMMi in the following.

TMMi is based on TMM itself, which in turn is based on the Capability Maturity Model (CMM) and CMMI, and was first proposed in 1998 [20]. The latest version of TMMi specification as of this writing is 1.0 [17] which is prepared and published by the TMMi Foundation in 2012.

Fig. 1 shows TMMi maturity levels and process areas and Fig. 2 shows its structure and components. As the structure outlines, each maturity level has several process areas (PA), and each process area has several specific goals and specific practices. In total, under the four maturity levels (2, 3 and 4), the TMMi [17] specified 50specific goals (SG) and 188 specific practices (SP). For example, under the level 2 (managed), there are five process areas, e.g., PA 2.1 (test policy and strategy). This PA has three SGs: SG 1-establish a test policy, SG 2-establish a test strategy, and SG 3-establish test performance indicators. The above SG 1, in turn, has three SPs: SP 1.1-define test goals, SP 1.2-define test policy, and SP 1.3-distribute the test policy to stakeholders.

In this context, it is also important to discuss the general process for TMA/TPI. In a 1999 book, Koomen and Pol nicely summarize that process as shown in Fig. 3, which starts with obtaining awareness, i.e., pinpointing the need for TMA/TPI.

2.2. Multivocal literature reviews

While SLR and SM studies are valuable, researchers have reported that "*the results of a SLR or a SM study could provide an established body of knowledge, focusing only on research contributions*" [21]. Since these secondary studies do not include the "grey" literature (non-published, nor peer-reviewed sources of information), produced constantly in a very large scale by practitioners, those studies do not provide much insight into the "state of the practice". For a practical (practitioner-oriented) field such as SE, synthesizing and combing both the state-of-the art and –practice is very important. Unfortunately, it is a reality that a large majority of software practitioners do not publish in academic forums [22], and this means that the voice of the practitioners is limited if we do not consider grey literature in addition to academic literature in review studies.

2.2.1. MLRs in other fields

The term Multivocal Literature Review (MLR) was defined in the early 1990 s in other fields, e.g., in educational research [8], as SLR which includes both the academic (formal) and the grey (informal) literature. The main difference between an MLR and a SLR or a SM is the fact that, while SLRs and SMs use as input only academic peer-reviewed articles, in MLRs, grey literature, such as blogs, white papers and web-pages, is also considered as input