The Journal of Systems and Software 111 (2016) 1–33

Contents lists available at ScienceDirect

The Journal of Systems and Software

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

Software test process improvement approaches: A systematic literature review and an industrial case study

Wasif Afzal[a,b,*], Snehal Alone[c], Kerstin Glocksien[c], Richard Torkar[c]

[a] School of Innovation, Design and Engineering, Mälardalen University, Västerås, Sweden
[b] Department of Computer Science, Bahria University, Islamabad, Pakistan
[c] Chalmers University of Technology, University of Gothenburg, Sweden

ARTICLE INFO

Article history:
Received 19 September 2014
Revised 13 August 2015
Accepted 27 August 2015
Available online 10 September 2015

Keywords:
Software test process improvement
Systematic literature review
Case study

ABSTRACT

Software test process improvement (STPI) approaches are frameworks that guide software development organizations to improve their software testing process. We have identified existing STPI approaches and their characteristics (such as completeness of development, availability of information and assessment instruments, and domain limitations of the approaches) using a systematic literature review (SLR). Furthermore, two selected approaches (TPI NEXT and TMMi) are evaluated with respect to their content and assessment results in industry. As a result of this study, we have identified 18 STPI approaches and their characteristics. A detailed comparison of the content of TPI NEXT and TMMi is done. We found that many of the STPI approaches do not provide sufficient information or the approaches do not include assessment instruments. This makes it difficult to apply many approaches in industry. Greater similarities were found between TPI NEXT and TMMi and fewer differences. We conclude that numerous STPI approaches are available but not all are generally applicable for industry. One major difference between available approaches is their model representation. Even though the applied approaches generally show strong similarities, differences in the assessment results arise due to their different model representations.

1. Introduction

It is a well-known fact that software testing is a resource-consuming activity. Studies show that testing constitutes more than 50% of the overall costs of software development (Harrold, 2000); and with the increasing complexity of software, the proportion of testing costs will continue to rise unless more effective ways of testing are found. One main focus of investigation in industry, for reducing cycle time and development costs, and at the same time increasing software quality, is improving their software testing processes (Collofello et al., 1996). However, state of practice in testing is sometimes ignored or unknown in software development organizations as testing is done in an *ad hoc* way (Bertolino, 2007) without designated testing roles being defined.

In the past, several software test process improvement (STPI) approaches have been developed to help organizations in assessing and improving their testing processes. To improve software testing process of a specific organization, an appropriate approach has to be found which suits their specific needs and the methodologies. Obviously, the expectations of the companies differ depending on, e.g., internal goals, maturity awareness and process knowledge. In our understanding, there is a need of consolidating available STPI approaches, along with their specific characteristics, in order to assist organizations in selecting the most appropriate approach.

This paper has an overall goal: to support industry in finding appropriate STPI approaches that fulfill the specific needs of an organization. This goal is fulfilled by two objectives: (1) to identify and evaluate existing STPI approaches and (2) to assist organizations in selecting and comparing the most appropriate STPI approaches. First, a general evaluation is applied to all approaches found by a systematic literature review (SLR). Second, a more specific and detailed evaluation is performed on two approaches using an industrial case study. The first part starts by finding a set of STPI approaches available in literature. Then these approaches are evaluated by a set of criteria. Besides providing information about the identified STPI approaches useful for further research, this evaluation constitutes the basis for the selection of approaches for the second part, i.e., the industrial case study. The second part starts with a pre-selection of applicable approaches based on the results of the first evaluation. A presentation of the pre-selected approaches and results of a voting scheme at

* Corresponding author at: School of Innovation, Design and Engineering, Mälardalen University, Västerås, Sweden. Tel.: +0046 732708813.
E-mail addresses: wasif.afzal@gmail.com, wasif.afzal@mdh.se (W. Afzal), Snehal.Alone@gmail.com (S. Alone), Kerstin.Glocksien@gmail.com (K. Glocksien), richard.torkar@cse.gu.se (R. Torkar).

http://dx.doi.org/10.1016/j.jss.2015.08.048
0164-1212/

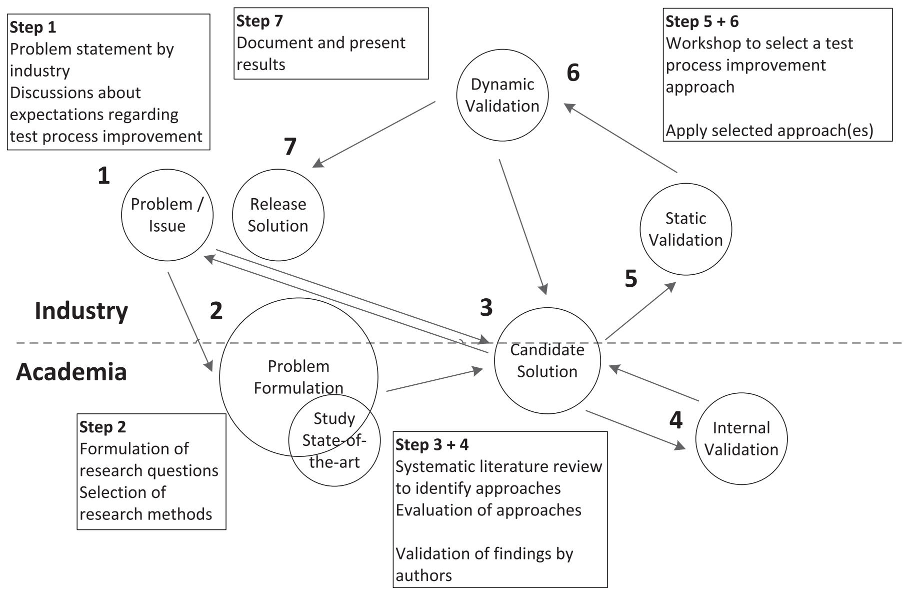

Fig. 1. Technology Transfer Model (originally published in Gorschek et al. (2006)).

the organization resulted in two approaches which are then applied in parallel at the organization. The selected approaches are examined and evaluated in more detail regarding their specific content. Finally, after application of both approaches at the organization, their results have been compared.

The rest of the paper is organized as follows: The next Section 2 describes the overall design of this paper. Section 3 presents the related work. Section 4 discusses the design of the SLR including the research questions, search strategy, study selection and quality assessment, data extraction, evaluation criteria for approaches and validation of results. Section 5 outlines the results of the SLR including the characteristics of 18 STPI approaches and listing approaches that are generally applicable in industry. Section 6 discusses the design of the case study while Section 7 discusses the case study results. The outcomes of this paper are discussed in Section 8 while the validity threats are discussed in Section 9. The major conclusions from this study appear in Section 10.

2. Overall study design

The design of this study is based on a model for technology transfer between academia and industry known as the Technology Transfer Model (Gorschek et al., 2006). The underlying theme of this model is that mutual cooperation is beneficial for both academia and industry. Researchers receive the opportunity to study industry relevant issues and validate their research results in a real setting. Practitioners, on the other hand, receive first-hand knowledge about new technology which helps them in optimizing their processes. A graphical overview of our study design based on the Technology Transfer Model is shown in Fig. 1 which has been adapted to the specific needs of our industrial partner.

The different steps in the design of this study based on the Technology Transfer Model are described as follows:

Step 1 – Problem/issue. A problem statement given by industry and discussions with company representatives about expectations and needs identify the problem as the unavailability of sufficient knowledge about the practiced testing process and a potential for process improvements.

Step 2 – Problem formulation. A preliminary study of the problem indicates the availability of software test process improvement (STPI) approaches providing frameworks and models to assess the current state of a testing process and to identify improvement suggestions. Based on this knowledge and industrial needs, the research questions along with appropriate research methodologies are identified.

Step 3 – Candidate solution. A systematic literature review (SLR) is conducted to identify available STPI approaches. The characteristics of these approaches are identified and an exclusion process provides a selection of generally applicable STPI approaches.

Step 4 – Internal validation. The findings from the SLR are partly validated by a number of authors of the primary studies identified by the SLR.

Step 5 – Static validation. The preselected generally applicable STPI approaches are presented in industry. The $100 method, a cumulative voting method (Rinkevics and Torkar, 2013), is used to select approaches to be applied in the organization.

Step 6 – Dynamic validation. The selected STPI approaches are applied in the organization. To assess the testing process, interviews are conducted and the data is analyzed based on the instructions given by the STPI approaches. Afterwards, the assessment results are compared based on a prior mapping of the content of the approaches.

Step 7 – Release solution. The results of the study are collected, documented and being presented in academia and industry.

Based on this overall design we decided to conduct the study by using two research methods, a systematic literature review (SLR) and a case study. The SLR covers Steps 3 and 4 of the model, candidate solutions and their characteristics are identified by the SLR and the results are internally validated. Steps 5 and 6 of the model, the static and dynamic validation, are explicitly covered by the case study. Moreover, we present in Table 1 the research goal, objectives, associated research questions, research method(s) used and relevant sections of the paper.

3. Related work

Software process improvement (SPI) frameworks involve assessment and improvement of software development processes. The need for such frameworks is motivated by the assumption that quality of a product is dependent on the process used to develop it. There

Table 1
Overall goal, objectives, research questions, research method and relevant section numbers.

Overall goal: To support industry in finding appropriate STPI approaches that fulfill the specific needs of an organization			
Objectives	Research questions (given in Sections 4.1 and 6)	Research method	Answered in
(1) To identify and evaluate existing STPI approaches	RQ 1	SLR	Section 5.1
	RQ 2	SLR	Section 5.2
	RQ 3	SLR	Section 5.3
(2) To assist organizations in selecting and comparing the most appropriate STPI approaches	$RQ_{cs}1$	Case study	Section 6.2
	$RQ_{cs}1.1$	SLR, Case study	Section 6.5
	$RQ_{cs}2$	Case study	Section 7.1
	$RQ_{cs}3$	Case study	Section 7.2

are several different SPI initiatives that are popular in industry. Card (2004) identifies them as Capability Maturity Model – Integrated, Six Sigma, Lean Development and ISO Standard 9001.

A common approach of many SPI frameworks is that actual processes are compared with best practices and any improvements are identified. This is referred to as the top-down approach (Thomas and McGarry, 1994). This is in contrast to a bottom-up approach where process change initiates based on knowledge of the organization and not based on prescribed best practices. The experience factory (Basili, 1993) is one example of bottom-up approach to SPI. Another distinction among SPI initiatives is with respect to their model architecture. A popular architecture is the staged/continuous representation of CMMi where improvements are judged with respect to capability and maturity levels. Another architecture is proposed by standards such as ISO 9001 that sets out requirements of a quality management system.

Available literature reviews in the area of SPI focus on the state of art in SPI (Zil-e-Huma et al., 2012), SPI applied in small and medium enterprises, both, in general (Pino et al., 2008), in a specific domain like web development (Sulayman and Mendes, 2011), and assessment of the impact of different SPI initiatives (Unterkalmsteiner et al., 2012).

Several case studies have been conducted with respect to CMM. The longitudinal study by Fitzgerald and O'Kane (1999) reports how a company achieved the CMM maturity levels in a period of four years. The case studies presented in Dangle et al. (2005) and Ilyas and Malik (2003) focus on the process changes needed to evolve from CMM level 2 to level 3 and to adapt company's existing processes to the processes proposed by CMM level 2. Experiences in actually performing the CMM assessment with regards to a specific process are reported in Kiiskila (1998). Comparison of multiple SPI approaches is given in Varkoi and Makinen (1998) and Wang et al. (1999). CMM and SPICE assessments are applied in two related software development units in Varkoi and Makinen (1998). The structures of both models are analyzed and a mapping between both models is performed for a specific process area. Finally, the assessed SPICE process capabilities and CMM maturity levels are compared. In Wang et al. (1999), a comparison of the assessment results, the robustness and the average assessment time of SPICE, CMM, BOOTSTRAP, ISO 9000, and SPRM is given.

Since the existing SPI frameworks (including CMM and CMMi) only provide limited attention to software testing (TMMi Foundation, 2010), the software testing community has created a number of its own improvement models. In some cases, these STPI models are complementary to SPI models since they are structured in a similar way. According to Garcia et al. (2014), about half of existing STPI approaches have a structure similar to CMM/CMMi.

Other STPI approaches are applicable in a different context such as PDCA-based software testing improvement framework (Xu-Xiang and Wen-Ning, 2010) is applicable when test processes are carried out as services by third party testing centers. Some STPI approaches use a bottom-up approach in the sense that they rely on identification of testing issues in the organization and then propose solutions. Observing practice (Taipale and Smolander, 2006) is one such example. The SLR part of this study presents these STPI approaches with respect to different characteristics.

There exists a literature study on software test process models by Garcia et al. (2014). They present a classification of 23 test process models based on model source, domain and publication year. They conclude that many of test process models are adapted or extended from TMMi and TPI. They also found a trend towards specialization of models to specific domains, such as automotive and embedded systems. We consider the SLR part of this study to be complementing Garcia et al. (2014) study whereby different related characteristics of STPI approaches are identified in much more detail.

Comparisons of STPI approaches have been reported in Swinkels (2000) and Farooq and Dumke (2008) but they are not complete with respect to reporting of all existing approaches. Swinkels (2000) compared the Testing Maturity Model (TMM) with the Test Improvement Model (TIM) and the Test Process Improvement Model (TPIM). With respect to comparison with TIM, TMM was found to be more comprehensive and detailed. In contrast, TPI checklist was found to be more detailed than TMM questionnaire. TMM also did not cover a number of TPI key areas. TMM was also found to be lacking in adequate guidelines on many process improvement issues when compared with TPI in Farooq and Dumke (2008).

4. Systematic literature review (SLR)

The first part of this paper identifies a comprehensive set of available STPI approaches using a SLR. We followed the guidelines for conducting a SLR as proposed by Kitchenham and Charters (2007). SLR provides a mechanism for evaluating, identifying and interpreting "all available research relevant to a particular research question, topic, area or phenomenon of interest" (Kitchenham and Charters, 2007). It summarizes the existing evidence concerning a technology.

4.1. Research questions

With the goal of identifying the existing STPI approaches, the following RQs are answered by the SLR:

RQ1: Which different STPI approaches can be found in literature?
RQ2: What are the specific characteristics of these STPI approaches?
RQ3: Which approaches are generally applicable in industry?

4.2. Data sources and search strategy

The search strategy was decided after conducting a pilot search using the search string "Software AND Testing AND Process AND Improvement" in all databases (ACM Digital Library, IEEE Xplore Digital Library, ScienceDirect and Springer Link). The search was restricted to title, abstract and keywords (and modified if required for

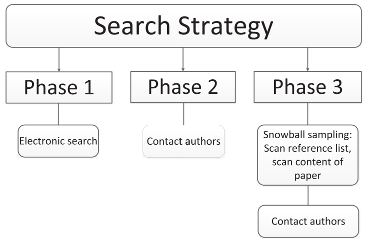

Fig. 2. Phases of the search strategy.

any database). The pilot search resulted in huge number of hits for Springer Link as it did not provide the same restriction options as other databases. After analyzing the results of the pilot search, a new search term "Software Testing Process" was identified. Using this phrase, further search terms were found from the titles of the papers found. The search terms were further complemented by words from relevant papers already known to us and by identifying synonyms for terms used in the titles of the found papers. The search terms were used with quotation marks for searching exact phrases. The final set of search terms used is following:

Software Testing Process, Software Test Process, Testing Process Improvement, Test Process Improvement, Test Maturity Model, Testing Maturity Model, Testing Process Model, Test Process Model, Software Testing Standard, Software Testing Optimization, Test Improvement Model, Testing Improvement Model

The search was divided into three phases (see Fig. 2). Fig. 3 further shows a complete picture of how final set of studies was reached.

Phase 1. In the first phase, we searched electronic databases. There was no limitation set on the publication year. We searched in the following databases:

- ACM Digital Library,
- IEEE Xplore Digital Library,
- ScienceDirect and
- Springer Link.

In Springer Link a limitation to search only in 'title', 'abstract' and 'keywords' was not possible, therefore we searched in full-text while for all other databases we searched in 'title', 'abstract' and 'keywords'. Table 2 outlines the numeric results of electronic search.

Phase 2. After getting the first data set, we performed the second phase of the search to have a more representative set of studies (see Fig. 3). In this phase, we contacted the authors of 22 candidate studies found in the electronic search of the first phase. The motive was to ask them of any papers that we might have missed from the first phase. The contact was established using the email addresses mentioned in the candidate studies or by email addresses found on the Internet. A total of 34 authors were contacted. For two authors no email addresses were available. Out of these 34 sent emails, 11 were undeliverable due to expired email addresses. We got a response from eight authors, out of which four provided relevant information.

Phase 3. In the third phase, snowball sampling (Goodman, 1961) was conducted. One researcher scanned the reference list of 16 studies to identify further papers (see Fig. 3). A second researcher also scanned the content of the studies to identify reference papers within the text that dealt with STPI approaches. These two ways of searching complemented each other since the titles of some papers in the reference lists did not always clearly indicate that the paper is dealing with STPI approaches; whereas for these references the relevance regarding the STPI research area was clearly indicated in the content of the paper. The number of found papers by snowball sampling is shown in Table 3.

Additionally, the third phase was completed by contacting the authors of the candidate studies identified by snowball sampling that dealt with previously unknown STPI approaches. Authors of three papers were contacted by email; in the end a total of five authors were contacted this way. Out of these five sent emails, four were not deliverable due to expired email addresses. One author replied but did not provide us with further research papers. After the conclusion of phase 3, we found a total of 35 papers after duplicates removal.

4.3. Study selection

For selecting the primary studies, the following inclusion criteria were applied, i.e., we included studies for which any of these questions were answered with 'yes':

- Does the paper talk about STPI approaches?
- Does the paper contain a case study on STPI?
- Does the paper contain a comparison between STPI approaches?
- Does the paper include an assessment done in any company on STPI?

Additionally, the following exclusion criteria were applied, i.e., we excluded papers that:

- only relate to software process improvement in general, not STPI in particular and,
- describe general software testing models.

The electronic database search (*phase 1*) resulted in a total of 404 papers. After eliminating duplicates, the number of papers reduced to 396 (see Table 2). The exclusion was done in several steps. Every step of the exclusion process was first performed by two researches independently.

4.3.1. Title and abstract exclusion

Two researchers independently conducted an inclusion and exclusion process by reading titles and abstracts, resulting in one of the three possible remarks for each paper – 'yes' (for inclusion) or 'maybe' (for further investigation in the next study selection step) and 'no' (for exclusion due to irrelevance to the research question). In this first step, the researchers agreed to exclude 320 papers.

To be able to measure the reliability of the inclusion and exclusion process the inter-rater agreement was calculated using Cohen's kappa coefficient (Cohen, 1960). The coefficient indicates the degree of agreement between two judges that exceeds the expected agreement by chance. Higher the value, more reliable are the results of the judgement as it can be expected that the judgement is reasonably based on knowledge and not on chance. The number of observed agreement was 354 (89.39% of the observations). The number of agreements expected by chance was 301.3 (76.08% of the observations). The Cohen's kappa result was 0.557. The strength of this agreement is considered to be 'moderate'. It is significantly higher than the expected agreement by chance and therefore a reliable judgement.

4.3.2. Introduction and conclusion exclusion

The researchers applied the detailed inclusion and exclusion criteria to the remaining 76 papers by reading 'introduction' and 'conclusion' sections, following the same process as in the previous step with three possible remarks for each paper.

The researchers agreed to exclude 38 papers and to include 16 papers. For 22 of the papers a discussion about inclusion or exclusion was required to resolve disagreements. The number of observed agreements was 57 (75.00% of the observations). The number of agreements expected by chance was 35.9 (47.23% of the observations). The Cohen's kappa result was 0.526. The strength of this agreement is 'moderate' and therefore considered as a reliable judgement.

After discussion, further 11 papers were excluded. The number of papers left after applying the detailed exclusion criteria was 27.

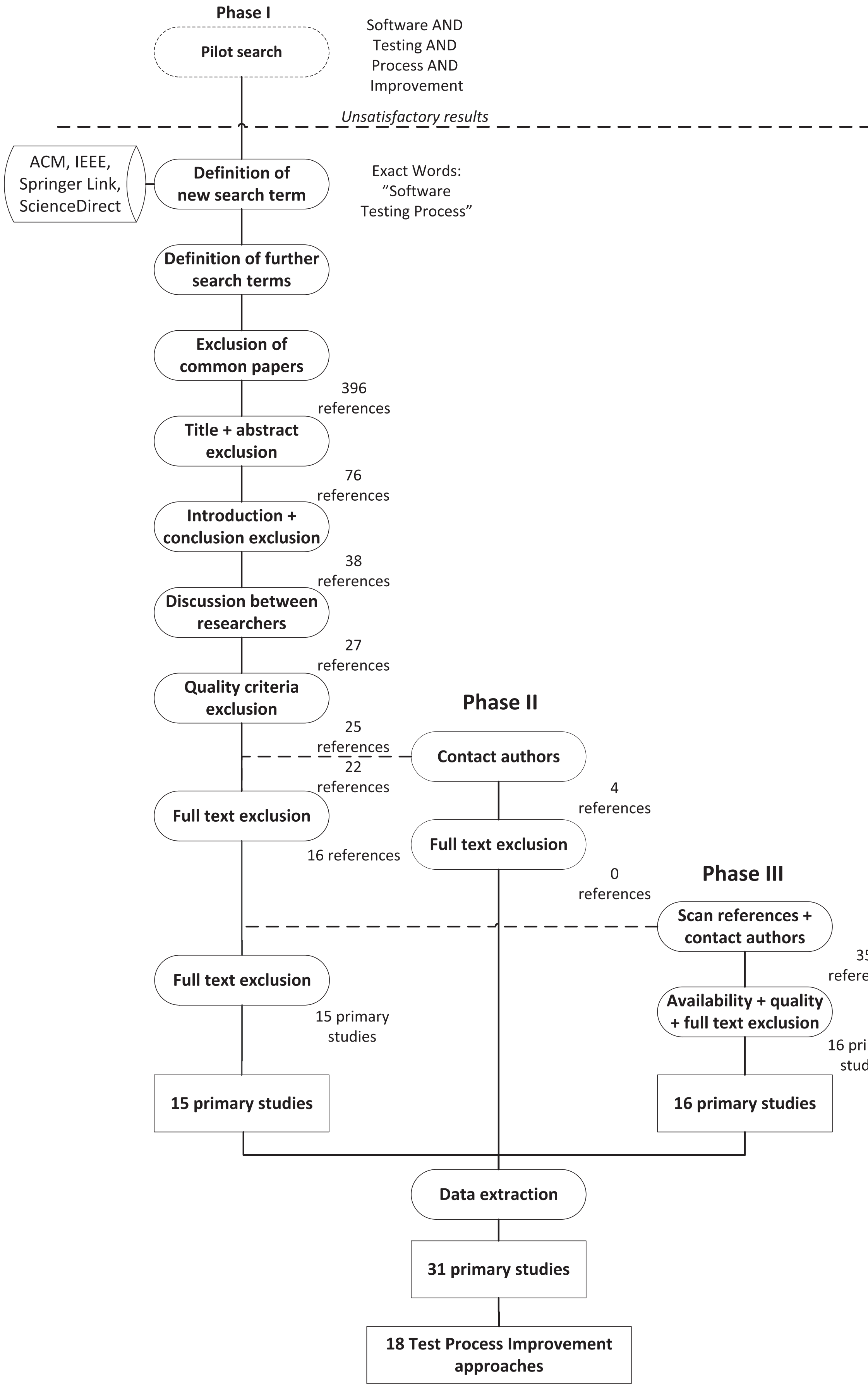

Fig. 3. Study selection process.

Table 2
Numeric results of electronic search.

Search term	ACM	ScienceDirect	IEEE	Springer Link
Software Testing Process	42	10	81	131
Software Test Process	21	1	28	132
Testing Process Improvement	2	1	5	39
Test Process Improvement	13	1	9	40
Testing Maturity Model	4	0	7	17
Test Maturity Model	5	0	1	17
Software Test Optimization	1	0	0	1
Test Process Model	5	0	12	32
Testing Process Model	3	0	7	32
Test Improvement Model	2	0	0	6
Testing Improvement Model	0	0	0	6
Software Testing Standard	3	0	1	8
Total per database (before duplicate removal)	101	13	151	461
Total per database (after duplicate removal)	74	12	129	187
Total (before duplicate removal)	404			
Total (after duplicate removal)	396			

Table 3
Numeric results of snowball sampling.

Original reference	Researcher A	Researcher B	Total after duplicate removal
Ryu et al. (2008)	3	3	3
Taipale and Smolander (2006)	1	1	1
Farooq et al. (2008)	5	5	6
Jung (2009)	10	10	10
Rana and Ahmad (2005)	0	0	0
Saldaña Ramos et al. (2012)	9	6	9
Burnstein et al. (1996)	2	1	2
Xu-Xiang and Wen-Ning (2010)	0	0	0
Tayamanon et al. (2011)	3	3	3
Jacobs and Trienekens (2002)	6	6	6
Kasoju et al. (2013)	1	0	1
Kasurinen et al. (2011)	8	6	8
Heiskanen et al. (2012)	9	9	9
Farooq and Dumke (2008)	8	8	8
Rasking (2011)	3	2	3
Reid (2012)	0	0	0

4.3.3. *Quality criteria exclusion*

Two papers were excluded by the application of the quality criteria described in Section 4.4.

4.3.4. *Exclusion in the context of contacting authors*

After applying the quality criteria, *Phase 2* of the search strategy – contacting authors – was started in parallel to Phase 1. During preparation for Phase 2 further three papers were excluded by consensus due to the irrelevance to the research topic.

4.3.5. *Full text exclusion*

At the end of Phase 1, the full-text of the remaining 22 papers was read and a further 6 papers were excluded by consensus. The remaining 16 papers identified as relevant to the topic were further considered as basis for conducting *Phase 3* – snowball sampling. Finally, we agreed to exclude one more paper based on re-reading the full-text.

The detailed exclusion process of Phase 1 of the search strategy resulted in 15 primary studies. Phase 2 of the search strategy, emailing the authors, resulted in four additional papers suggested by them, but these were later excluded when applying the exclusion criteria. In Phase 3 of the search strategy, 35 references found by snowball sampling were further investigated. Out of these 35 candidate studies, 12 papers were not freely available and 5 were excluded by reading the full-text. A further three papers were excluded based on the criteria specified for quality assessment (see Section 4.4).

In conclusion, the 15 primary studies found by the electronic database search were complemented by 16 primary studies found by snowball sampling.

4.4. *Study quality assessment*

We did not restrict studies based on a specific research method, so both quantitative and qualitative studies were considered. We did not devise multiple study quality assessment criteria but used a simple criterion that if a paper is assessed not to be peer-reviewed, it was excluded. Two papers, Meng (2009) and Xin-ke and Xiao-Hui (2009), were excluded as major parts were found to be identical. During the snowball sampling, one paper was excluded because it was not written in English and two references were excluded because they were not peer reviewed papers.

4.5. *Data extraction*

The data extraction was divided into two phases: (1) identifying STPI approaches described by the primary studies (RQ1) and (2) extracting detailed information about the approaches (RQ2). During the first phase, the name and, if available, the abbreviation of the STPI approach presented in the investigated paper was extracted.

For the second phase a data extraction form was prepared. For each STPI approach identified in the first phase of data extraction, the following information was extracted: 'Based on/influenced by', 'Domain', 'Developed by', 'Status of development', 'Completeness of information', 'Assessment model', 'Assessment procedure', 'Assessment instrument', 'Improvement suggestions', 'Process reference model', 'Maturity structure', 'Model representation', 'Character of approach', 'Structure/components', 'Addressing' and 'Process areas'.

The extracted characteristics of the approaches can be explained as follows:

Based on/influenced by: Earlier developed models or frameworks that function as basis or that have influenced the development of this approach.

Domain: A specific domain which this approach is addressing. If empty, a specific domain is either not mentioned or it is explicitly said that the approach is universally applicable.

Developed by: An institute, foundation or cooperation that developed the approach. If empty, the approach was developed by a single researcher or a smaller group of researchers, and an institute, foundation or cooperation was not explicitly mentioned.

Status of development: There are two possible dimensions of the status of development: ‘under development’ or ‘completed’. If the approach was validated by case studies, surveys or experiments, this is also mentioned.

Completeness of information: There are three dimensions regarding the completeness of the information possible: ‘concept’, ‘brief description’ or ‘detailed description’. Papers assessed as ‘concept’ only present the idea of the approach. Normally, approaches that are assessed as ‘under development’ are only presented as concepts in the respective study. For approaches with ‘detailed descriptions’, all the information is available to apply the approach. Detailed information about the assessment process, the components and the structure of the approach is available. ‘Brief descriptions’ provide more information than concepts but not all elements of the approach are described in detail.

Assessment model: An assessment model provides a framework/structure for the results of the assessment. The assessment results might be maturity levels that determine the state of practice of the assessed organization.

Assessment procedure: It is checked if the approach provides instructions how to perform the assessment.

Assessment instrument: It is checked if the approach provides an instrument, e.g., a questionnaire, which is used for the assessment.

Improvement suggestions: It is checked if the approach provides information about processes that need improvement to be able to move to a higher assessment/maturity level.

Process reference model: It is checked if the approach provides a reference model representing the ideal process which the organizations should be aiming for to reach the highest level of maturity.

Maturity structure: It is checked if the approach uses maturity levels to assess an organization’s test process. If yes, the maturity levels are listed.

Model representation: Two possible types of model representations are considered: ‘continuous’ or ‘staged’. In a continuous representation, each process area has a number of maturity levels, so that the maturity level of each process area can be assessed and improved individually. In a staged representation, a maturity level is composed of a set of specific process areas. To reach a higher maturity level, all requirements of all the process areas in that and the preceding maturity levels (if any) have to be satisfied.

Character of approach: There are two dimensions, ‘qualitative’ or ‘quantitative’. Qualitative approaches investigate the test process based on qualitative data, e.g., through interviews with employees. Quantitative approaches use quantitative data like metrics for the assessment of the test process.

Structure/components: Describes the structure of the approach and its components.

Addressing: If the approach is addressing specific roles in the organization, these are listed here.

Process areas: Lists the aspects of the testing process that are investigated by the approach.

4.6. Evaluation criteria

In order to examine if the STPI approaches are generally applicable in industry, the following evaluation criteria were devised:

- Has the development of the approach been completed?
- Is broad information about the approach available? (Completeness of information is more than a brief description.)
- Is there an assessment instrument (e.g., a questionnaire) available for this approach?
- Is the approach not specific to a domain?

STPI approaches, for which one or more of these questions were answered with ‘no’, were considered not generally applicable in industry (an exception to this rule was made for approaches where an assessment instrument was missing but with status of development being complete and presence of detailed description were still considered as being generally applicable).

It is to be noted that this evaluation criteria should be used with caution for STPI approaches that are domain-specific. Since our investigation is meant to find generally applicable STPI approaches in industry, one of our evaluation criteria excludes domain-specific approaches. There is a chance that a company in a particular domain (e.g. automotive) may still find a domain-specific approach most appropriate. Therefore, a company needs to make a decision regarding an approach to use by keeping in view their specific context.

4.7. Validation of results

The findings of the SLR were validated by the feedback from a set of authors of the selected primary studies. We contacted the authors by email (who had replied during *Phase 2* of the search strategy) as well as the authors of the studies identified by the snowball sampling. A total of seven authors were contacted. Three authors replied and gave feedback on our evaluation. With one author we conducted an interview in which he answered our validation questions.

We provided them with a list of all STPI approaches that we found in the SLR and asked them if this list is complete. Furthermore, we presented them our evaluation criteria for finding generally applicable approaches in industry and the particular inclusion/exclusion result for the particular approach presented by the contacted author. Individually, the authors were asked if they agree to the evaluation of their approach.

One of the authors of Taipale and Smolander (2006) validated our list of approaches as “good” and stated that it even contain approaches unknown to him. One of the authors of Heiskanen et al. (2012) agreed to our evaluation regarding the ATG add-on for TPI. He stated that an spreadsheet had been used for the assessment which had not been published.

In a short interview about TMMi conducted with the author of Rasking (2011) he agreed to our evaluation results for TMMi and also confirmed the results of our systematic review as “very complete”.

5. Review results

In the following section, the research questions are answered with the help of SLR findings.

5.1. Which different STPI approaches can be found in literature?

The STPI approaches found by the literature review are presented in Table 4. In total, 18 approaches have been identified. The studies Farooq and Dumke (2008), Farooq and Dumke (2007), Kulkarni (2006) and Swinkels (2000) have been identified as primary studies related to the research question since they are discussing STPI approaches. However, these studies are not listed in Table 4 because they are not explicitly presenting one specific approach but rather comparing several approaches.

Table 4
Found approaches.

Ref	Approach	Abbreviation
Ryu et al. (2008)	Ministry of National Defense-Testing Maturity Model	MND-TMM
Taipale and Smolander (2006)	Observing practice	–
Farooq et al. (2008)	Meta-measurement approach	–
Jung (2009)	Embedded Test Process Improvement Model	Emb-TPI
Rana and Ahmad (2005), Burnstein et al. (1996), Tayamanon et al. (2011), Burnstein (2003), Burnstein et al. (1999), Jacobs et al. (2000), Suwannasart (1996), Homyen (1998)	Testing Maturity Model	TMM
Xu-Xiang and Wen-Ning (2010)	Plan-Do-Check-Action (PDCA)-based software testing improvement framework	–
Jacobs and Trienekens (2002)	Metrics Based Verification and Validation Maturity Model	MB-VV-MM
Kasoju et al. (2013)	Evidence-based Software Engineering	–
Kasurinen et al. (2011)	Self-Assessment framework for ISO/IEC 29119 based on TIM	–
Heiskanen et al. (2012)	Test Process Improvement Model for Automated Test Generation	ATG add-on for TPI
Reid (2012)	Software Testing Standard ISO/IEC 29119, ISO/IEC 33603	–
Ericson et al. (1997)	Test Improvement Model	TIM
Karlström et al. (2005)	Minimal test practice framework	MTPF
Koomen and Pol (1999), Koomen (2002)	Test Process Improvement	TPI
TPI (2004)	TPI® Automotive	TPI® Automotive
v. Ewijk et al. (2013)	TPI® NEXT	TPI® NEXT
Rasking (2011), van Veenendal (2008)	Test Maturity Model integration	TMMi®
Steiner et al. (2012); Schweigert et al. (2014)	Test SPICE	–

5.2. What are the specific characteristics of these STPI approaches?

We have grouped the STPI approaches in to four categories:

- TMM and related approaches.
- TPI and related approaches.
- Standards and related approaches.
- Individual approaches.

Before describing these approaches in detail, Fig. 4 shows the dependencies between the different STPI approaches and their relation to test process models, standards, reference models, process improvement approaches, etc., which influenced their development. The following paragraphs describe the identified approaches. Brief background information and the most important characteristics are pointed out for each approach.

5.2.1. TMM and related approaches

5.2.1.1. TMM – Testing Maturity Model. The Testing Maturity Model was developed by a research group at the Illinois Institute of Technology in the late 1990s. Its purpose is to assess and improve testing processes in software development organizations. Furthermore it can be used as a model to represent the ideal incrementally growing testing process. Especially assessments from inside the company are possible. Amongst other sources, the development of TMM was influenced by CMM. The need for the development of the model emerged since existing evaluation frameworks did not address testing in a sufficient way. The structure of TMM is also inspired by CMM. It represents a staged model and consists of the following components:

- Five maturity levels: Initial, Definition, Integration, Management and Measurement, Optimization/Defect Prevention and Quality Control.
- Maturity goals (MG), maturity subgoals (MSG) and activities and tasks with responsibilities (ATR).
- An assessment model (TMM-AM).

The characteristics of TMM are given in Table 5.

5.2.1.2. TMMi® - Test maturity model integration. TMMi is generally known as the successor of TMM. It was developed by the TMMi Foundation, a non-profit organization, founded in 2005 by a group of leading test and quality practitioners. Their aim was to develop a testing model which covers the experience and best practices of a broad group of experts and would find acceptance in industry. Besides TMM as a development basis, TMMi was influenced by CMMi. TMMi consists of:

- Five maturity levels: Initial, Managed, Defined, Measured, Optimization.
- Process areas in each maturity level.
- Required components: Specific and generic goals.
- Expected components: Specific and generic practices.
- Informative components: Sub-practices, example work products, notes, examples or references.

The TMMi maturity levels have been inspired by the TMM maturity structure but further developed according to industry needs. The introduction of required, expected and informative components was established due to the influence of CMMi. Most generic goals and practices were even adopted from CMMi.

The characteristics of TMMi are given in Table 6.

5.2.1.3. MND-TMM – Ministry of National Defense-Testing Maturity Model. MND-TMM was developed to address the specific needs of weapon software system development. It combines the concepts of several approaches. It was influenced by TMM and TMMi and uses the continuous representation of CMMi. Furthermore, an OWL ontology is used to describe the elements of the model. Most elements of MND-TMM have been adopted from TMMi like specific and generic goals.

The model consists of ten process areas which are summarized in four categories – military, process, infrastructure and techniques. Each process area has five maturity levels. Due to the use of a continuous model the maturity of each process area can be assessed individually.

The characteristics of MND-TMM are given in Table 7.

5.2.1.4. MB-VV-MM – Metrics based verification and validation maturity model. The MB-VV-MM is a quantitative framework to improve validation and verification processes. Metrics are used to select process improvements and to track and control the implementation of improvement actions. The approach was based on TMM and enhanced by additions to specially support the validation and verification process. Similar to TMM, it consists of five maturity levels.

The characteristics of MB-VV-MM are given in Table 8.

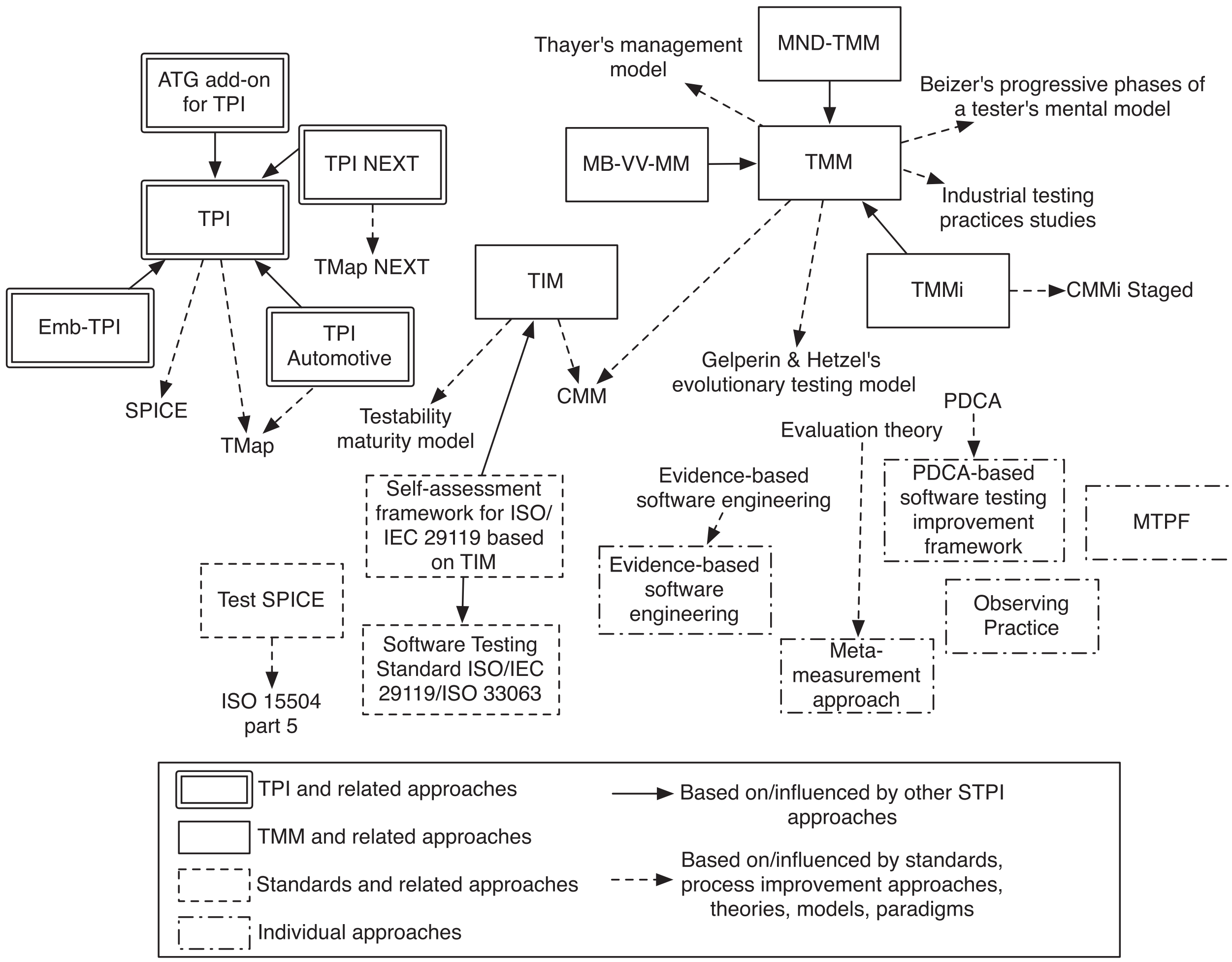

Fig. 4. Dependencies of STPI approaches.

Table 5
Characteristics of TMM.

Characteristics	
Approach	TMM – Testing Maturity Model
Reference	Rana and Ahmad (2005), Burnstein et al. (1996), Tayamanon et al. (2011), Burnstein et al. (1999), Homyen (1998), Burnstein (2003), Jacobs et al. (2000), Suwannasart (1996)
Based on/influenced by	CMM, Gelperin and Hetzel's evolutionary testing model, Industrial testing practices studies, Beizer's progressive phases of a tester's mental model, Thayer's management model
Domain	–
Developed by	Illinois Institute of Technology, USA
Status of development	Complete, Validated in an experiment
Completeness of information	Detailed description, Additional information: team selection and training
Assessment model	Yes
Assessment procedure	Available
Assessment instrument	Available, Questionnaire, Mainly yes/no questions + open questions, Individual interviews after first round of pre-defined questions
Improvement suggestions	Available, Recommendation of testing tools and test-related metrics
Process reference model	No
Maturity structure	Yes – 1: Initial, 2: Phase-Definition, 3: Integration, 4: Management and Measurement, 5: Optimizing/Defect prevention and quality control
Model representation	Staged
Character of approach	Qualitative
Structure/components	Maturity levels, Maturity goals (MG), Maturity subgoals (MSG), Activities, tasks, and responsibilities (ATR), Metrics, Tool recommendations, Critical views (managers, developers, users/clients)
Addressing	Test managers, Test groups, Software quality assurance staff
Process areas	Testing and debugging goals and policies, Test planning process, Testing techniques and methods, Test organization, Technical training program, Software life cycle, Controlling and monitoring, Review Test measurement program, Software quality evaluation, Defect prevention, Quality control, Test process optimization

Table 6
Characteristics of TMMi®.

Characteristics	
Approach	TMMi® – Test Maturity Model integration
Reference	Rasking (2011), van Veenendal (2008)
Based on/influenced by	CMMi (staged representation), TMM
Domain	–
Developed by	TMMi Foundation
Status of development	Complete
Completeness of information	Detailed description
Assessment model	Yes
Assessment procedure	Available
Assessment instrument	Not available
Improvement suggestions	Available
Process reference model	No
Maturity structure	Yes – 1: Initial, 2: Managed, 3: Defined, 4: Measured, 5: Optimization
Model representation	Staged
Character of approach	Qualitative
Structure/components	Maturity levels, Process areas, Specific goals, Specific practices, Generic goals, Generic practices
Addressing	Test managers, Test engineers, Software quality professionals
Process areas	Test policy and strategy, Test planning, Test monitoring and control, Test design and execution, Test environment, Test organization, Test training program, Test lifecycle and integration, Non-functional testing, Peer reviews, Test measurement, Product quality evaluation, Advanced reviews, Defect prevention, Quality control, Test process optimization

Table 7
Characteristics of MND-TMM.

Characteristics	
Approach	MND-TMM – Ministry of National Defense-Testing Maturity Model
Reference	Ryu et al. (2008)
Based on/influenced by	TMM
Domain	Defense – military weapon systems
Developed by	Partially supported by Defense Acquisition Program Administration and Agency for Defense Development
Status of development	Under development
Completeness of information	Concept
Assessment model	Yes
Assessment procedure	Not available
Assessment instrument	Not available
Improvement suggestions	Not available
Process reference model	No
Maturity structure	Yes – 5 levels
Model representation	Staged + continuous, Similar to the continuous approach of CMMi
Character of approach	Qualitative
Structure/components	Maturity levels, Categories, Test process areas (TPAs), Specific goals, Specific practices, Sub practices, Generic goals, Common features
Addressing	–
Process areas	Military: Software quality evaluation, Process: Test strategy Test planning, Test process management, Infrastructure: Test organization, Test environment, Testware management, Techniques: Testing techniques, Test specification, Fault management

5.2.1.5. TIM – Test Improvement Model. The Test Improvement Model serves as a guidebook for improvements of the test process and focuses explicitly on cost-effectiveness and risk management. Its intention is to identify the current state of practice with strong and weak elements and to make suggestions how to further strengthen the strong elements and to improve the weak elements. It was inspired by SEI's Capability Maturity Model and Gelperin's Testability Maturity Model.

TIM belongs to the group of continuous models and it is seen as the first step of the PDCA method, the planning phase. The model consists of five key areas. Each key area has five levels of maturity: initial, baselining, cost-effectiveness, risk-lowering and optimizing, which are each represented by one overall goal and several subgoals.

The characteristics of TIM are given in Table 9.

5.2.2. TPI and related approaches

5.2.2.1. TPI – Test Process Improvement. The Test Process Improvement Model was developed in a Dutch company called IQUIP in the late 1990s. The model is based on the test approach TMap. It helps analyzing the current situation and identifying strengths and weaknesses of an organization's test process.

TPI is a continuous approach. It consists of 20 key areas which represent different points of view on the test process. Each key area can have up to four levels of maturity. Checkpoints are used to determine the maturity level of each key area. They are requirements that have to be met for a test process to be classified in a specific level of maturity.

A Test Maturity Matrix provides an overview of the testing maturity of the assessed organization by highlighting the satisfied checkpoints and maturity levels per key area.

The characteristics of TPI are given in Table 10.

5.2.2.2. TPI® NEXT. TPI NEXT is the successor of TPI, developed by the Dutch company Sogeti (a corporate merger of IQUIP and other companies). Compared to the original TPI approach the number of key areas in TPI NEXT has been reduced to 16 and additional elements—enablers and clusters—have been introduced to the model to more efficiently address industry needs in Test Process Improvement.

The characteristics of TPI® NEXT are given in Table 11.

5.2.2.3. TPI automotive. A further approach developed by the Dutch company Sogeti is TPI automotive. It follows the same principles as

Table 8
Characteristics of MB-VV-MM.

Characteristics	
Approach	MB-VV-MM – Metrics Based Verification and Validation Maturity Model
Reference	Jacobs and Trienekens (2002)
Based on/influenced by	TMM
Domain	–
Developed by	Consortium of industrial companies (defense and civil systems, telecommunication and satellites, consumer and professional electronics), consultancy and service agencies (software quality, testing, and related vocational training) and an academic institute (Frits Philips Institute, University of Technology – Eindhoven), Netherlands
Status of development	Under development, Validated in various experiments
Completeness of information	Concept
Assessment model	Yes
Assessment procedure	Not available
Assessment instrument	Not available
Improvement suggestions	Not available
Process reference model	No
Maturity structure	Yes – 1: Initial, 2: Repeatable, 3: Defined, 4: Managed and aligned, 5: Optimizing
Model representation	Staged, Planned to address continuous aspects
Character of approach	Quantitative/qualitative
Structure/components	Maturity levels, Process areas, Process goals, Metrics, Generic practices
Addressing	–
Process areas	V&V Environment, V&V Design methodology, V&V Monitor and control, V&V Project planning, V&V Policy and goals, Peer reviews, V&V Lifecycle embedding, Training and program, Organization embedding, Qualitative process measurement, Quality measurement and evaluation, Organizational alignment, Process optimization, Quality management, Defect prevention

Table 9
Characteristics of TIM.

Characteristics	
Approach	TIM – Test Improvement Model
Reference	Ericson et al. (1997)
Based on/influenced by	CMM, TMM – Testability Maturity Model
Domain	–
Developed by	–
Status of development	Complete
Completeness of information	Brief description
Assessment model	Yes
Assessment procedure	Not available
Assessment instrument	Not available, No use of yes/no-questions
Improvement suggestions	Not available
Process reference model	No
Maturity structure	Yes – Initial, Baselining, Cost-effectiveness, Risk-lowering, Optimizing
Model representation	Unknown
Character of approach	Qualitative
Structure/components	Key areas, Maturity levels, Overall goal for the level, Subgoals, Activities, Checkpoints
Addressing	–
Process areas	Organization, Planning and tracking, Test cases, Testware, Reviews

Table 10
Characteristics of TPI.

Characteristics	
Approach	TPI – Test Process Improvement
Reference	Koomen and Pol (1999), Koomen (2002)
Based on/influenced by	SPICE, TMap
Domain	–
Developed by	Sogeti
Status of development	Complete
Completeness of information	Detailed description
Assessment model	Yes
Assessment procedure	Available
Assessment instrument	Available, Checkpoints
Improvement suggestions	Available
Process reference model	Yes
Maturity structure	Yes – Controlled, Efficient, Optimized
Model representation	Continuous
Character of approach	Qualitative
Structure/components	Key areas (20), Maturity levels, Checkpoints (300), Test maturity matrix, Improvement suggestions, Dependencies between different levels of the key areas
Addressing	–
Process areas	Test strategy, Life-cycle model, Moment of involvement, Estimation and planning, Test specification techniques, Static test techniques, Metrics, Test tools, Test environment, Office environment, Commitment and motivation, Test functions and training, Scope of methodology, Communication, Reporting, Defect management, Testware management, Test process management, Evaluation, Low-level testing

Table 11
Characteristics of TPI® NEXT.

Characteristics	
Approach	TPI® NEXT
Reference	v. Ewijk et al. (2013)
Based on/influenced by	Tmap NEXT, TPI
Domain	–
Developed by	Sogeti
Status of development	Complete
Completeness of information	Detailed description
Assessment model	Yes
Assessment procedure	Available
Assessment instrument	Available
Improvement suggestions	Available
Process reference model	Yes
Maturity structure	Yes
Model representation	Continuous
Character of approach	Qualitative
Structure/components	Key areas (16), Maturity levels, Checkpoints (157), Clusters, Enablers, Test maturity matrix, Improvement suggestions, Dependencies between different levels of the key areas
Addressing	–
Process areas	Stakeholder commitment, Degree of involvement, Test strategy, Test organization, Communication, Reporting, Test process management, Estimating and planning, Metrics, Defect management, Testware management, Methodology practice, Tester professionalism, Test case design, Test tools, Test environment

Table 12
Characteristics of TPI® Automotive.

Characteristics	
Approach	TPI®Automotive
Reference	TPI (2004)
Based on/influenced by	TMap, TPI
Domain	Automotive
Developed by	Sogeti, German automotive industry
Status of development	Complete
Completeness of information	Detailed description
Assessment model	Yes
Assessment procedure	Available
Assessment instrument	Available, Checkpoints
Improvement suggestions	Available
Process reference model	Yes
Maturity structure	Yes – maximum 4 levels (individual for each key area)
Model representation	Continuous
Character of approach	Qualitative
Structure/components	Key areas (21), Maturity levels, Checkpoints, Test maturity matrix, Improvement suggestions, Dependencies between different levels of the key areas
Addressing	–
Process areas	Test strategy, Life-cycle model, Moment of involvement, Estimation and planning, Test design techniques, Static test techniques, Metrics, Test automation, Test environment, Office and laboratory environment, Commitment and motivation, Test functions and training, Scope of methodology, Communication, Reporting, Defect management, Testware management, Test process management, Evaluation Low-level testing, Integration testing

TPI but was specifically adapted to the needs of software testing in automotive industry.

The characteristics of TPI automotive are given in Table 12.

5.2.2.4. ATG add-on for TPI – Test Process Improvement Model for Automated Test Generation. This approach represents an add-on for the existing TPI to address the aspects of automated test generation in Test Process Improvement, especially the use of formal methods. The add-on extends TPI by:

- new maturity levels in the key areas of ‘Static test techniques’ and ‘Test specification techniques’,
- new key areas ‘Modeling approach’, ‘Use of models’, ‘Test confidence’, ‘Technological and methodological knowledge’ and
- new checkpoints.

The characteristics of ATG add-on for TPI are given in Table 13.

5.2.2.5. Emb-TPI – Embedded Test Process Improvement Model. Embedded TPI focuses on improving the testing process for embedded software by especially considering hardware issues of testing. The model consists of the following elements:

- capability model,
- maturity model,
- test evaluation checklist,
- evaluation and improvement procedure and,
- enhanced test evaluation model.

The characteristics of Emb-TPI are given in Table 14.

5.2.3. Standards and related approaches

5.2.3.1. Test SPICE. The intention of developing Test SPICE was to provide a process reference model (PRM) and process assessment model (PAM) specific for test process assessment in conformance with the requirements of ISO/IEC 15504 II. Using ISO/IEC 15504 V as a starting

Table 13
Characteristics of ATG add-on for TPI.

Characteristics	
Approach	ATG add-on for TPI – Test Process Improvement Model for Automated Test Generation
Reference	Heiskanen et al. (2012)
Based on/influenced by	TPI
Domain	Automated testing
Developed by	–
Status of development	Complete, Validated in a case study
Completeness of information	Brief description
Assessment model	Yes
Assessment procedure	Not available
Assessment instrument	Available, Checkpoints
Improvement suggestions	Not available
Process reference model	No
Maturity structure	Yes – Maximum 4 levels (individual for each key area)
Model representation	Continuous
Character of approach	Qualitative
Structure/components	Key areas, Maturity levels, Checkpoints, Test maturity matrix, Improvement suggestions, Dependencies between different levels of the key areas
Addressing	–
Process areas	Test strategy, Life-cycle model, Moment of involvement, Estimation and planning, Test specification techniques, Static test techniques, Metrics, Test tools, Test environment, Office environment, Commitment and motivation, Test functions and training, Scope of methodology, Communication, Reporting, Defect management, Testware management, Test process management, Evaluation, Low-level testing, Modeling approach, Use of models, Test confidence, Technological and methodological knowledge

Table 14
Characteristics of Emb-TPI.

Characteristics	
Approach	Emb-TPI – Embedded Test Process Improvement Model
Reference	Jung (2009)
Based on/influenced by	TPI
Domain	Embedded software
Developed by	–
Status of development	Complete, Validated in a case study and a survey
Completeness of information	Brief description
Assessment model	Yes
Assessment procedure	Not available
Assessment instrument	Not available
Improvement suggestions	Not available
Process reference model	No
Maturity structure	Yes
Model representation	Continuous
Character of approach	Qualitative
Structure/components	Key areas, Maturity levels, Checkpoints, Test maturity matrix, Improvement suggestions, Dependencies between different levels of the key areas
Addressing	–
Process areas	18 key areas with 6 categories: Test process, Test technique, Test automation, Test quality, Test organization, Test infrastructure

point and reusing its structure, the Test SPICE model was developed by:

- identically transferring processes from ISO/IEC 15504 V to Test SPICE,
- replacing original processes from ISO/IEC 15504 V with specific test processes,
- renaming processes of ISO/IEC 15504 V and,
- inserting new specific test processes to Test SPICE.

Currently Test SPICE V3.0 is in the final phase of the international review process (Schweigert et al., 2014). Test SPICE V3.0 focusses on rearrangement of the relationship to ISO/IEC 15504 V, alignment to ISO 29119-2 and more attention to technical testing processes, e.g. test automation and test data management (Schweigert et al., 2014).

The characteristics of Test SPICE are given in Table 15.

5.2.3.2. Software testing standard ISO/IEC 29119, ISO/IEC 33063. ISO/IEC 29119 is a testing standard. The need for this standard was identified due to the traditionally poor coverage of testing in standards. Available standards with respect to testing cover only small, particular parts of testing, not the overall testing process.

ISO/IEC 29119 is divided into five parts: concepts and definitions, test processes, test documentation, test techniques and keyword driven testing. By working in accordance with the process proposed in the standard, a specific product quality can be guaranteed. In addition, ISO/IEC 33063, the process assessment standard related to the testing standard, provides a means to assess the compliance of a testing process to ISO/IEC 29119.

The characteristics of software testing standard ISO/IEC 29119 /ISO 33063 are given in Table 16.

5.2.3.3. Self-assessment framework for ISO/IEC 29119 based on TIM. The goal of this approach is to provide an assessment framework that checks the compliance of an organization's test process with the standard ISO/IEC 29119. Therefore, the concept of the Test Improvement Model (TIM) with its maturity levels has been combined with the propositions of the standard. The model is divided into three levels: organizational, project and execution level. Similar to TIM, this approach has five maturity levels: initial, baseline, cost-effectiveness,

Table 15
Characteristics of Test SPICE.

Characteristics	
Approach	Test SPICE
Reference	Steiner et al. (2012) and Schweigert et al. (2014)
Based on/influenced by	ISO 15504 part 5
Domain	–
Developed by	SQS Group
Status of development	Complete
Completeness of information	Detailed description
Assessment model	Yes
Assessment procedure	Not available
Assessment instrument	Not available
Improvement suggestions	Not available
Process reference model	Yes
Maturity structure	No
Model representation	–
Character of approach	Qualitative
Structure/components	Process categories, Process groups, Processes
Addressing	–
Process areas	Process categories and groups: Primary life cycle processes, Test service acquisition, Test service supply, Test environment operation, Testing Supporting life cycle processes, Test process support, Organizational life cycle processes, Management Resource and infrastructure, Process improvement for test, Regression and reuse engineering

Table 16
Characteristics of software testing standard ISO/IEC 29119 /ISO 33063.

Characteristics	
Approach	Software testing standard ISO/IEC 29119 /ISO 33063
Reference	Reid (2012)
Based on/influenced by	–
Domain	–
Developed by	ISO/IEC
Status of development	Under development
Completeness of information	Brief description
Assessment model	Yes
Assessment procedure	Not available
Assessment instrument	Not available
Improvement suggestions	Not available
Process reference model	Yes
Maturity structure	No
Model representation	–
Character of approach	Qualitative
Structure/components	Process descriptions, Test documentation, Test techniques
Addressing	–
Process areas	Test policy, Organizational test strategy, Test plan, Test status report, Test completion report, Test design specification, Test case specification, Test procedure specification, Test data requirements, Test environment requirements, Test data readiness report, Test environment readiness report, Test execution log, Incident report

risk-lowering and optimizing, and also follows the continuous approach which means that the key areas are assessed separately.

The characteristics of self-assessment framework for ISO/IEC 29119 based on TIM are given in Table 17.

5.2.4. Individual approaches

5.2.4.1. Meta-measurement approach. This approach focuses on the specification and evaluation of quality aspects of the test process. It is based on the concept of Evaluation Theory (Ares Casal et al., 1998) and it has been adapted to address the test process sufficiently. It consists of the following steps:

- Target (software test processes).
- Evaluation criteria (quality attributes).
- Reference standard (process measurement profiles).
- Assessment techniques (test process measurements).
- Synthesis techniques (quality matrix, quality indexes).
- Evaluation process.

The characteristics of meta-measurement approach are given in Table 18.

5.2.4.2. Plan-do-check-action (PDCA)-based software testing improvement framework. The PDCA-based software testing improvement framework was developed to specifically address test processes provided as services by third party testing centers. The concept of this approach is based on the hypothesis that knowledge management plays an important role in process improvements. The framework is divided into the following phases:

- Build a learning organization through knowledge management.
- Plan the adaptive testing processes.
- Plan implementation and data analysis.
- Continuous improvement.

The characteristics of PDCA-based software testing improvement framework are given in Table 19.

5.2.4.3. Evidence-based software engineering. In this individual approach, improvements for the test process are identified by the use of evidence-based software engineering. First, challenges in the testing process of an organization are identified by interviews. Then, solutions to these challenges are searched by a systematic literature

Table 17
Characteristics of self-assessment framework for ISO/IEC 29119 based on TIM.

Characteristics	
Approach	Self-assessment framework for ISO/IEC 29119 based on TIM
Reference	Kasurinen et al. (2011)
Based on/influenced by	ISO/IEC 29119, TIM
Domain	-
Developed by	Supported by the ESPA-project
Status of development	Complete, Validated in pilot study with pre-existing data (four different case organizations)
Completeness of information	Brief description
Assessment model	Yes
Assessment procedure	Available
Assessment instrument	Available, Open questions
Improvement suggestions	Not available (only individual examples from the case study)
Process reference model	Yes
Maturity structure	Yes – 0: Initial, 1: Baseline, 2: Cost-effectiveness, 3: Risk-lowering, 4: Optimization
Model representation	Continuous
Character of approach	Qualitative
Structure/components	Processes, Maturity levels
Addressing	Software designer, Software architect, Manager, Test manager, Project leader, Tester
Process areas	Organizational test process (OTP), Test management process (TMP), Test planning process (TPP), Test monitoring and control process (TMCP), Test completion process (TCP), Static test process (STP), Dynamic test process (DTP)

Table 18
Characteristics of meta-measurement approach.

Characteristics	
Approach	Meta-measurement approach
Reference	Farooq et al. (2008)
Based on/influenced by	Evaluation theory
Domain	–
Developed by	–
Status of development	Under development
Completeness of information	Concept
Assessment model	Yes
Assessment procedure	Not available
Assessment instrument	Not available
Improvement suggestions	Not available
Process reference model	No
Maturity structure	No
Model representation	–
Character of approach	Quantitative
Structure/components	Target, Evaluation criteria, Reference standard, Assessment techniques, Synthesis techniques, Evaluation process
Addressing	–
Process areas	Activities, Product (document, test cases, etc.), Resource (software, hardware, personnel), Roles

Table 19
Characteristics of PDCA-based software testing improvement framework.

Characteristics	
Approach	PDCA-based software testing improvement framework
Reference	Xu-Xiang and Wen-Ning (2010)
Based on/influenced by	PDCA
Domain	Third party testing center
Developed by	–
Status of development	Complete (thesis work)
Completeness of information	Brief description
Assessment model	No
Assessment procedure	Not available
Assessment instrument	Not available
Improvement suggestions	Not available
Process reference model	No
Maturity structure	No
Model representation	–
Character of approach	Unknown
Structure/components	Test improvement framework divided into phases: Plan, Do, Check, Action
Addressing	–
Process areas	–

Table 20
Characteristics of evidence-based software engineering.

Characteristics	
Approach	Evidence-based Software Engineering
Reference	Kasoju et al. (2013)
Based on/influenced by	Evidence-basedsoftware engineering
Domain	Automotive software (applied in this domain, but not necessarily limited to it)
Developed by	–
Status of development	Complete
Completeness of information	Brief description
Assessment model	No
Assessment procedure	Not available
Assessment instrument	Not available
Improvement suggestions	Not available (only individual examples from the case study)
Process reference model	No
Maturity structure	No
Model representation	–
Character of approach	Qualitative
Structure/components	Multi-staged evidence-based software engineering research process, Case study with interviews to identify strengths and weaknesses of the testing process, Domain specific literature review/mapping to find solutions to identified problems, Value stream mapping identify process wastes, show locations of improvements
Addressing	–
Process areas	–

Table 21
Characteristics of observing practice.

Characteristics	
Approach	Observing practice
Reference	Taipale and Smolander (2006)
Based on/influenced by	–
Domain	Software products and applications of an advanced technical level, mission critical, real-time-environments (applied in this domain, but not necessarily limited to it)
Developed by	Supported by the ANTI-project
Status of development	Complete, Factors affecting testing know-how and organizations have not been addressed yet, Validated in a case study with 4 organizational units
Completeness of information	Detailed description
Assessment model	No
Assessment procedure	Not available
Assessment instrument	Available, structured and semi-structured questions, 4 theme-based interview rounds
Improvement suggestions	Not available (only individual examples from the case study)
Process reference model	No
Maturity structure	No
Model representation	–
Character of approach	Qualitative
Structure/components	Interviews, Grounded theory to analyze data, Classify data into categories, Illustrate interdependencies of the categories with cause-effect graphs, Process improvement propositions
Addressing	Managers of development, Managers of testing, Testers, System analyst
Process areas	Factors affecting testing, for example: Involvement of testing in the development process, Management of the complexity of testing, Risk-based testing, Communication and interaction between development and testing, Use and testing of software components, Adjusting testing according to the business orientation of an organization's unit, Factors affecting testing know-how and organization, Categories derived from data analysis: Involvement of testing in the development process, Testing schedules, Communication and interaction between development and testing, Planning of testing, Use of software components, Complexity of testing

review. Finally, an improved test process is presented by value-stream mapping.

The characteristics of evidence-based software engineering are given in Table 20.

5.2.4.4. Observing practice. In this approach the test process is studied by conducting detailed interviews with varying roles involved in testing in several interview rounds. The data gained by the interviews is analyzed by the use of grounded theory. Problems and at the same time possible solutions are identified by the analysis.

The characteristics of observing practice are given in Table 21.

5.2.4.5. MTPF – Minimal test practice framework. MTPF is a lightweight approach which addresses smaller organizations. Its goal is to increase acceptance of proposed improvements by the involvement of the entire organization. The framework addresses five categories which correspond to areas in testing. The introduction of process improvement is leveled in three phases which are adapted to the size of the organization.

The characteristics of MTPF are given in Table 22.

To allow for a side by side comparison of different STPI approaches, Table 23 presents a condensed summary of relevant characteristics of these approaches (some characteristics such as 'Structure/Components', 'Process areas', 'Developed by' and 'Addressing' are omitted in this condensed summary due to space limitations). Fig. 5 presents the timelines of the different STPI approaches, based on the first appearance of an approach (year of initial publication), follow-up publications, successor approaches and references from studies or related work. We combine the timeline with information regarding status of development and completeness of information.

Table 22
Characteristics of MTPF.

Characteristics	
Approach	MTPF – Minimal test practice framework
Reference	Karlström et al. (2005)
Based on/influenced by	–
Domain	–
Developed by	–
Status of development	Complete, Validated in a case study and a survey
Completeness of information	Brief description
Assessment model	No
Assessment procedure	Not available
Assessment instrument	Not available
Improvement suggestions	Not available
Process reference model	No
Maturity structure	No
Model representation	–
Character of approach	Qualitative
Structure/components	3 phases depending on the size of the organizational unit, Introduction phase consisting of 5 steps: prepare, introduce, review, perform, evaluate
Addressing	–
Process areas	Problem and experience reporting, Roles and organization issues, Verification and validation, Test administration, Test planning

5.3. Which approaches are generally applicable in industry?

To answer this question, the evaluation criteria specified in Section 4.6 were applied on the 18 STPI approaches identified by the SLR. This evaluation procedure led to a set of six approaches being generally applicable. These six approaches are TMM, TMMi, TPI, TPI NEXT, Test SPICE and observing practice. The application of evaluation criteria is given in Table 24 where the six generally applicable approaches are highlighted in bold.

Even though TPI NEXT is the successor of TPI, and the concept of TMMi is based on TMM and TMMi is often also seen as the successor of TMM, these approaches are still considered separately in this paper.

6. Case study

The second part of this paper is a case study where we evaluate two selected STPI approaches with respect to their content and assessment results. The guidelines for conducting and reporting case study research given in Runeson and Höst (2009) are used as a basis for completing this case study.

The objective of our case study was to identify STPI approaches valuable for the case organization, apply them and compare their content and assessment results. Robson (2002) call such objectives as *exploratory* since they seek to understand what is happening in little-understood situations, to seek new insights and to generate ideas for future research. Moreover, based on the insights gained from conducting the SLR and case study, we reflect on the information needs of an organization to select appropriate STPI approaches.

In order to fulfill our objective, the following research questions were formulated:

- $RQ_{cs}1$: Which approaches are valuable for test process improvements in the company under study?
- $RQ_{cs}1.1$: What typical information is required by an organization to select appropriate STPI approaches?
- $RQ_{cs}2$: How well can the content of the selected approaches be mapped to each other for an effective assessment in our case organization?
  To be able to effectively compare the assessment results of STPI approaches applied in the case organization, the similarities and differences with respect to content of the selected approaches need to be identified. Besides being an important input for $RQ_{cs}3$, and thus affects the case study, the answers to $RQ_{cs}2$ provide significant information in regards to a general evaluation of the applied STPI approaches.
- $RQ_{cs}3$: How do the results of the selected approaches differ after applying them?

Since individuals and their processes significantly influence the answers to our research questions (i.e. the context is multidisciplinary), therefore case study was considered a better choice over e.g. action research. Moreover, the assessment of STPI approaches in an industrial setting has an observational character, thus further indicating the applicability of a case study. Action research is also conducted in a natural setting but compared to case studies, the researcher is directly involved in the process of improvement or change intended by research. The process of research itself influences outcome of the study. Since research questions $RQ_{cs}1$ and $RQ_{cs}3$ only have an observational character and do not require actual process changes within the case organization initiated by researchers, case study was preferred over action research.

The elements of the case study design are summarized in Table 25.

6.1. Case description and context

The organization under study is a part of Volvo IT which is a subsidiary of the Volvo Group, a large Swedish automotive organization. The team develops and maintains information systems within the product development (PD) and purchasing (PU) area for an external customer.

Both areas, PD and PU, consist of several different information systems and applications developed in a number of different programming languages. Systems in the PD area are handling product data needed for product development in automotive industry. PU systems manage, for example, suppliers information. In total, 45 employees are working in the case organization, of which 20 are located in Gothenburg (Sweden) and 25 in Bangalore (India).

Apart from line management, the following roles could be found within the organization: maintenance manager, project manager, coordinator, system analyst, business analyst and developer. Smaller teams consisting of system and/or business analysts and developers are responsible for one or several of the systems/applications in either the PD or PU area. The developers are mainly located in India.

Testing is not seen as a major activity of the development or maintenance process. Within the team, there are no designated testing roles. Even though a corporate test policy is available for Volvo IT, it is unknown to what extent these guidelines are followed by the team.

Table 23
A condensed summary of the characteristics of different STPI approaches (under 'Model representation', the letters S and C stand for Staged and Continuous respectively; ✓ and × are analogous to Yes and No respectively).

	TMM	TMMi	MND-TMM	MB-VV-MM	TIM	TPI	TPI NEXT	TPI Automotive	ATG Add-on for TPI	Emb-TPI	Test SPICE	ISO/IEC 29119 / ISO 33063	Self-assess. framework	Meta-measure. approach	PDCA-based	Evidence-based	Observ. practice	MTPF
Domain	-	-	Defense	-	-	-	-	Automotive	Automated Testing	Emb. software	-	-	-	-	Third party testing center	-	-	-
Assessment model	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	×	×	×
Assessment procedure	✓	×	×	×	×	✓	✓	✓	×	×	×	×	✓	×	×	×	×	×
Assessment instrument	✓	×	×	×	×	✓	✓	✓	✓	×	×	×	✓	×	×	×	✓	×
Improvement suggestions	✓	✓	×	×	×	✓	✓	✓	×	×	×	×	×	×	×	×	×	×
Process reference model	×	×	×	×	×	✓	✓	✓	×	×	✓	✓	✓	×	×	×	×	×
Maturity structure	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	×	✓	×	×	×	×	×
Model representation	S	S	S / C	S	-	C	C	C	C	C	-	-	C	-	-	-	-	-
Character of approach	Qualitative	Qualitative	Qualitative	Quant./Qual.	Qualitative	Qualitative	Qualitative	Qualitative	Qualitative	Qualitative	Qualitative	Qualitative	Qualitative	Quantitative	-	Qualitative	Qualitative	Qualitative

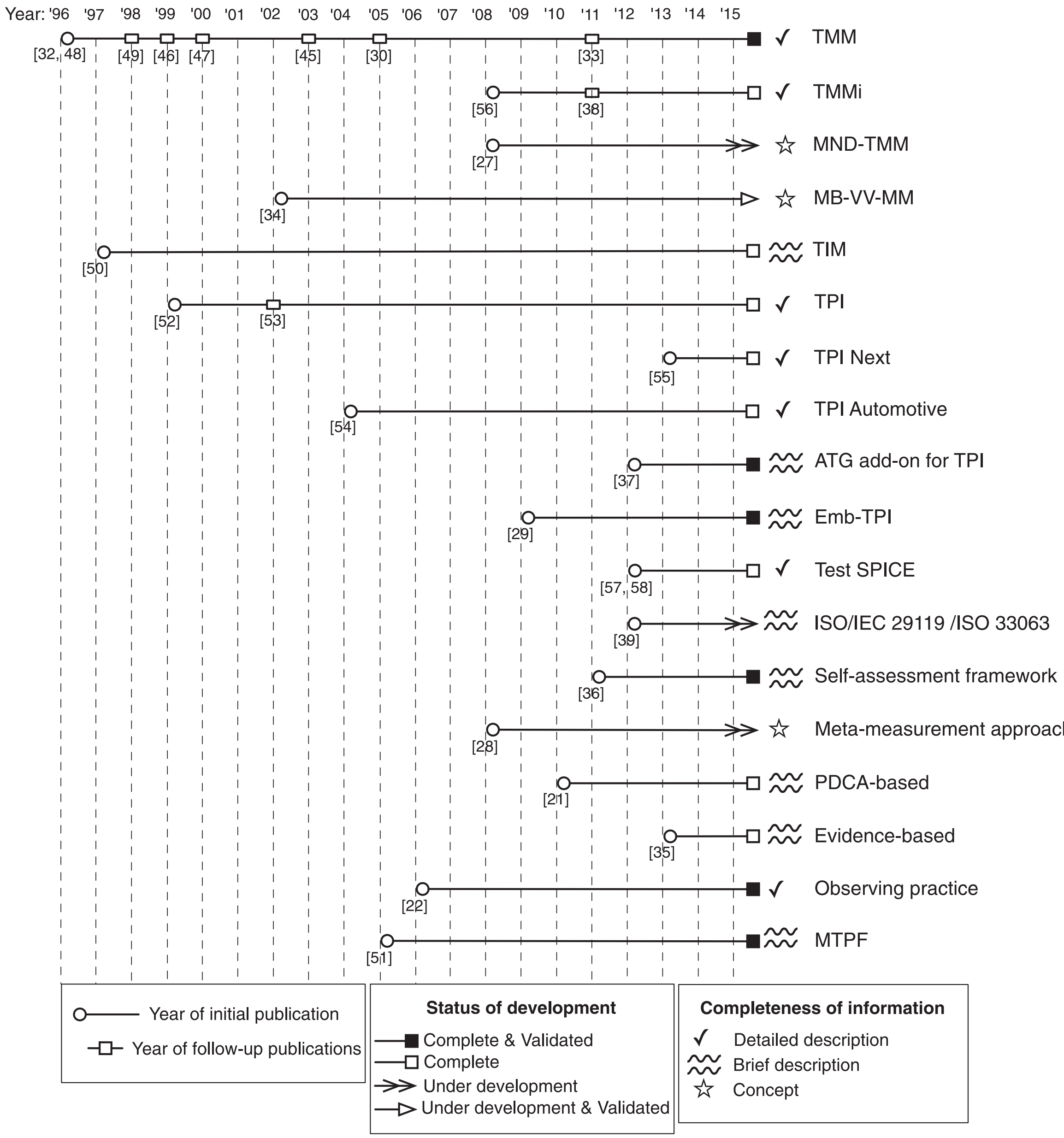

Fig. 5. Timelines of STPI approaches with additional information.

The processes are rather set in accordance to the requirements of the external customer. Moreover, it is perceived that each team member follows her own testing process.

However, there is a general consensus that the quality of development deliverables is good. This notion is mainly based on the lack of frequent or serious complaints from customer side.

The testing policy is provided by a globally operating department of Volvo IT, called ADT (Application Development and Technology). The department is responsible for establishing standard processes in different areas of software development. Furthermore, they offer the service of testing process assessment.

The study is conducted as a *holistic* case study (Yin, 2003) since the context is considered being the specific company where the team members involved in testing and their testing process are studied as a whole.

During the whole study, key personnel, called as 'organization representatives' in the following sections, supported us in decision making processes, e.g., interviewee selection. The 'organization representatives' were representing different levels of authority within the organization. They were line manager of the organization, the maintenance manager of each area, and one system/business analyst of each area.

6.2. Selection of STPI approaches for the case organization using a workshop

The answer to the RQ 3 of our SLR (Section 5.3) gave us a set of approaches that are generally applicable in industry. The selection of the actual approaches to be applied in the case organization was done during a workshop.

The participants of the workshop were the 'organization representatives' and two persons from outside the organization, who had shown interest in participating. Both of the external participants were members of the ADT team within Volvo IT (mentioned in Section 6.1). They worked in the area of testing in general and TPI in particular and had a keen interest in our study.

Table 24
Application of evaluation criteria to 18 STPI approaches. The six generally applicable approaches are highlighted in bold.

	Evaluation criteria			
	Development completed?	More than a brief description?	Availability of an instrument?	Not specific to a domain?
TMM	✓	✓	✓	✓
TMMi	✓	✓	×	✓
MND-TMM	×	×	×	×
MB-VV-MM	×	×	×	✓
TIM	✓	×	×	✓
TPI	✓	✓	✓	✓
TPI NEXT	✓	✓	✓	✓
TPI Automotive	✓	✓	✓	×
ATG add-on for TPI	✓	×	✓	×
Emb-TPI	✓	×	×	×
Test SPICE	✓	✓	×	✓
ISO/IEC 29119-ISO 33063	×	×	×	✓
Self-Assessment framework	✓	×	✓	✓
Meta Measurement	×	×	×	✓
PDCA-based	✓	×	×	×
Evidence-based	✓	×	×	✓
Observing practice	✓	✓	✓	✓
MTPF	✓	×	×	✓

Table 25
Case study design.

Study characteristics		
Objective	Exploratory	Identify STPI approaches valuable for the case organization, apply them and compare their content and their assessment results
Case	Holistic	Investigating the testing process and the team members involved in testing as a whole
Data collection	Qualitative	Collecting data through interviews, observation and documents
Triangulation	Data (source) triangulation	Interviews, observations and document analysis

The workshop consisted of two steps: a presentation held by two researchers followed by a cumulative voting.

6.2.1. Presentation

The presentation started with an introduction to the research process and the objective of conducting the workshop. The results of the SLR as well as the evaluation criteria used for the pre-selection of applicable STPI approaches were presented. Finally, the pre-selected approaches were explained in detail. The information provided for each approach was based on the following parameters:

Developed by Which company, organization, research group or individual researcher developed this approach?
Based on Which approach/methodology is the approach based on? For example, it might be based on CMMi.
Model representation Which type of model representation is used in the approach? Continuous or staged?
Key elements What are the key elements of the approach? For example, checkpoints or specific goals and practices.
Process areas Which areas are investigated by the approach? For example, test strategy, stakeholder commitment or test policy.
Assessment procedure What is the assessment procedure of the approach? For example, interviews with open-ended questions.

Thereafter, detailed content-wise examples of the investigated process areas were provided.

During the presentation of the characteristics of the pre-selected approaches and the content-wise examples, particular attention was given on emphasizing the differences between the approaches without rating them as advantages or disadvantages. The approaches were presented in an objective way without emphasizing any specific approach to prevent biased decisions.

After the presentation, printed material about each of the presented approaches was handed out to all participants and an open discussion about the approaches was held. The discussion phase was mainly used to answer questions regarding the presentation. The workshop finally ended with cumulative voting to decide which approach(es) should be applied in the organization under study.

6.2.2. Cumulative voting

The decision which STPI approach was to be applied in the case organization was done by using the $100 method (Leffingwell and Widrig, 2003).

The $100 method is a cumulative voting method to make a selection between several alternative options. Each participant of the voting is provided with a virtual $100 to distribute between the options. The participants can distribute any amount between 0 and 100 on any of the options. The only restriction is that each participant has to distribute $100 in total at the end of the voting. The higher an amount spent on an option the more priority that option has. The option with the highest result will be selected.

All participants of the workshop except for the researchers had a vote with equal weighting. Each participant's vote consisted of $100 which could be distributed arbitrarily between the presented approaches with any amount between 0 and 100.

The results of the voting are presented in Table 26.

Table 26 shows that TPI NEXT received the highest scores with 400 points and TMMi got the second highest scores with 279 points. Clearly behind are the scores for the third and fourth ranking. On the third rank is TMM with 20 points and observing practice reached the fourth rank with only 1 point. TPI and Test SPICE did not get any votes.

Considering the knowledge and experience in the field of test process improvement of two of the participants, the interpretation of the results requires a different perspective. Unlike the other participants of the workshop, participants 6 and 7 already had detailed knowledge about TPI. One of them even had experience in performing assessments using TPI.

Table 26
Results from applying cumulative voting (the straight line after Participant 5 separates the two group of participants).

	TPI	TPI®NEXT	TMM	TMMi®	TestSPICE	Observing practice
Participant 1	0	40	20	39	0	1
Participant 2	0	50	0	50	0	0
Participant 3	0	60	0	40	0	0
Participant 4	0	50	0	50	0	0
Participant 5	0	0	0	100	0	0
Participant 6	0	100	0	0	0	0
Participant 7	0	100	0	0	0	0
Total	0	400	20	279	0	1

Table 27
Interviewee description.

Interviewee no.	Role	Experience in organization [years]	Location	Area
1	System analyst	2	Gothenburg	PU
2	System analyst	2.5	Gothenburg	PD
3	System analyst	24	Gothenburg	PU
4	System analyst	10	Gothenburg	PD
5	Project manager	2	Gothenburg	PU/PD
6	Business analyst	22	Gothenburg	PD
7	Application developer	2	Bangalore	PD
8	Application developer	2.5	Bangalore	PU
9	Application developer	1.2	Bangalore	PU
10	Application developer	2.5	Bangalore	PU
11	Application developer	5	Bangalore	PU
12	Application developer	2	Bangalore	PD

If the votes of participants 6 and 7 were disregarded, TMMi would have received the highest scores with 279 points, compared to TPI NEXT with 200, TPI with 20 and observing practice with 1 point. Due to the fact that in both perspectives TPI NEXT and TMMi clearly obtained the highest rankings, we decided to apply these two approaches in the case study.

6.3. Data collection

The data needed for the case study (i.e. test process assessments) was mainly collected through interviews. Additionally, testing documents and processes that were identified during the interviews as relevant for the assessment, were studied and observed. The data from several sources was collected for triangulation purposes, to make our conclusions stronger and to eliminate effects of one interpretation of one single data source (Runeson and Höst, 2009).

6.3.1. Interviewee selection

The participants were selected with the help of the 'organization representatives'. The selection of a team member as an interviewee was based on her involvement in testing activities. Furthermore, it was required for the selected interviewees to be a representative sample of the population. Therefore, both areas, PD and PU, and both development sites, Gothenburg and Bangalore, were covered as well as all roles related to testing activities.

Two members from 'organization representatives' were also selected as interviewees. Besides their professional knowledge of the case organization's testing process, they were selected because their interviews served as pilot studies. An anonymized list of all interviewees stating their roles, working area and their current location is specified in Table 27.

6.3.2. Interview design

The interview questions were designed with respect to the aim of having joint interviews for both approaches. Due to this objective we decided to have semi-structured interviews with mainly open ended questions. This strategy aimed in getting maximum information from one question. With general phrased, open ended questions we aimed in combining the overall content of all key areas of TPI NEXT and all process areas of TMMi in one common questionnaire. Furthermore, available questionnaires from STPI approaches served as input to the process of interview question development (Taipale and Smolander, 2006; Kasoju et al., 2013). The feedback from the interviewees of the two pilot interviews was additionally used to reframe and rephrase the questions after conducting these first two interviews. The semi-structured interview approach allowed us to adjust the course of the interview, the set of asked questions and their level of detail according to the interviewees role and her knowledge.

The interviews were structured into following themes:

Introduction A short introduction to the research topic and process was given.
Warm-up questions Questions regarding the interviewee's age, educational background, years of experience in the case organization and in IT in general were covered in this theme.
Overview of work tasks Questions regarding the interviewee's usual work tasks and her involvement in testing.
Questions specific to testing This was the major section in which we tried to cover all process areas, such as regression testing, test environment, testing with respect to product risks, test plan, test cases, testing tools, defects and training on testing.
Statistical questions about the interview These questions were asked to get their opinion on interview design, questions, duration and the general feeling about the interview.

The complete set of pre-designed questions is given in Appendix A.

6.3.3. Execution of the interview

Prior to the interview phase, emails were sent to all interviewees briefly describing the purpose and relevance of the interviews. Except for the two pilot interviews, the duration of the interviews was set to a maximum of 60 min. All interviews were recorded in an audio format and, additionally, notes were taken. The interviews were conducted in person with the participants in Gothenburg (Sweden) while telephone interviews were conducted with the interviewees in Bangalore (India).

Table 28
An example of data analysis for TPI NEXT assessment.

Key area: Stakeholder commitment	
Checkpoint 1: The principal stakeholder is defined (not necessarily documented) and known to the testers.	Yes
Checkpoint 2: Budget for test resources is granted by and negotiable with the principal stakeholder.	No
Checkpoint 3: Stakeholders actually deliver the committed resources.	No
Checkpoint 4: The principal stakeholder is responsible for a documented product risk analysis (the input for the test strategy).	No
Checkpoint 5: All relevant stakeholder are defined (not necessarily documented) and known to the testers.	No
Checkpoint 6: Stakeholders actively acquire information on the quality of both the test process and the test object.	No
Checkpoint 7: The stakeholders proactively take action on aspects that affect the test process. This includes changes in the delivery sequence of the test object and changes in the project scope.	No
Checkpoint 8: Line management acknowledges that test process improvement comes with the need for increased learning time for which resources are provided.	No
Checkpoint 9: Stakeholders are willing to adapt their way of working to suit the test process. This includes the software development and requirements management.	No
Checkpoint 10: An adapted way of working by the stakeholder to suit demands of the test process is jointly evaluated by the test organization and the stakeholder.	No

As basis for the data analysis, the contents of all interviews were briefly transcribed after the interview phase. The individual transcript of each interview was sent to the respective interviewee with the request to check the content for its correctness.

6.3.4. Observation

Observation helps to understand processes better by seeing the actual execution. For few processes/system features, the researchers sat next to the interviewees when they were executing tests or performing a test-related process.

6.3.5. Document analysis

Process documents such as test policy, software test description, test cases, test plans, testing reports and all other documents related to testing were studied to gain a better understanding of the organizational processes and standards. This in turn helped in understanding and analyzing the interview data.

6.4. Data analysis procedures

The data collection phase was followed by data analysis. Since the main focus lay on assessment of state of practice with respect to test process and not the identification of improvements, the instructions regarding improvement suggestions were neglected during data analysis. Especially, the process of TPI NEXT was affected by this decision.

The main element of the assessment with TPI NEXT is the verification of the checkpoints provided by the model. Based on the interview data, the documents studied and the processes observed, Researcher A checked the fulfillment of the checkpoints for each key area. Since the default maturity level of any organization in TPI NEXT is 'initial', we started the assessment from the next maturity level of 'controlled'. Fulfilled checkpoints were marked with 'Yes' and not fulfilled checkpoints were marked with 'No'. The results were documented in a spreadsheet provided on the TPI NEXT website. The spreadsheet automatically produces the TPI NEXT Test Maturity Matrix which highlights the fulfilled checkpoints in the respective maturity level of each key area.[1] Due to the limitation to the assessment of the state of practice, the consideration of clusters was disregarded. As an example, the first key area in TPI NEXT assessment is 'stakeholder commitment'. This key area has a total of ten checkpoints, fulfillment of these will characterize its level of maturity. For our case organization, only one checkpoint in maturity level 'controlled' was fulfilled, represented with an answer 'Yes' in Table 28. This answer was given because there was evidence found in test artefacts in our case organization pointing to the fulfillment of this checkpoint. The other checkpoints were not fulfilled and are represented with answer 'No' in Table 28. The TPI NEXT Test Maturity Matrix, which is automatically generated, thus characterized the fulfillment degree of this key area as being low.

In a formal assessment of TMMi, the result is based on the degree of fulfillment of specific and generic goals. TMMi provides a rating scale which specify the degree of fulfillment in detail. In an informal assessment, as described by the TMMi Foundation, this procedure is not proposed. However, since we needed to build a basis on which we could compare the results of the TPI NEXT assessment and the TMMi assessment , we adapted the assessment procedure for this purpose. Based on the interview data, Researcher *B* checked the fulfillment of the specific and generic goals associated with the process areas of maturity Level 2. The fulfillment for each specific and generic practice was classified by the following rating: 'fully fulfilled', 'partly fulfilled' or 'not fulfilled'.

If the testing process is performed exactly like the practices proposed by TMMi or by an alternative, this practice is marked as 'fully fulfilled'. If only particular steps in the practices are fulfilled, this practice is marked as 'partly fulfilled'. If a TMMi practice is not followed at all, this practice is marked as 'not fulfilled'. Due to the staged character of the TMMi model, an assessment of a higher level is not needed if the goals of the preceding level are not fully fulfilled. Therefore only the process areas and goals of TMMi Level 2 were investigated. As an example, the process area 'test policy and strategy' has one of its specific goals as 'establish a test policy'. It has three specific practices, namely 'define test goals', 'define test policy' and 'distribute the test policy to stakeholders'. These specific practices were assessed being either 'fully fulfilled', 'partially fulfilled' or 'not fulfilled' based on the available evidence in our case organization (Table 29). For example, Table 29 shows that specific practice 'SP1.1: define test goals' was assessed as being 'not fulfilled' as there was no available evidence of defined test goals. The other specific practices only had partial fulfillment, for example for the specific practice: 'SP1.3: Distribute the test policy to stakeholders', the team members in our case organization were not aware of the test policy, although it was available on their web portal.

The assessment procedure of TPI NEXT and the informal assessment of TMMi do not require the assessor to provide particularly strong or multiple evidences for her decision if a checkpoint or a goal is fulfilled or not. Hence, the decision relies on the assessor's interpretation with respect to the compliance with the model. Both researchers agreed that a checkpoint or a goal was stated as fulfilled if an indication of the fulfillment was given by at least one interviewee.

6.5. Typical information needs of an organization for selecting STPI approaches

This section lists the typical information needs of an organization when selecting an STPI approach. These information needs are based

[1] http://www.tmap.net/en/tpi-NEXT/downloads

Table 29
An example of data analysis for TMMi assessment.

Process area PA2.1: Test policy and strategy		
Specific goal (SG)	Specific practices (SP)	Assessment result
SG1: Establish a test policy	SP1.1: Define test goals	not fulfilled
	SP1.2: Define test policy	partly fulfilled
	SP1.3: Distribute the test policy to stakeholders	partly fulfilled

on insights gained from selecting STPI approaches for our case organization and by conducting the SLR.

- Our SLR results (Section 5) have already indicated that there are a number of STPI approaches but most of them do not provide sufficient information. This makes them difficult to apply in practice. Therefore a pre-selection of approaches based a concrete evaluation criteria is needed. We present one such set of criteria in Section 4.6. This pre-selection not only helped our case organization to deal with a smaller subset but also helped them focus their selection efforts on complete approaches.
- As confirmed by experts working in STPI domain, they before did not know some of the approaches identified by our SLR. Therefore, an organization needs to disseminate information about STPI approaches through e.g., workshops. In our workshop (Section 6.2), we presented a condensed summary of pre-selected approaches that covered six important elements: Developed by, Based on, Model representation, Key elements, Process areas and Assessment procedure. This condensed summary was followed by detailed content-wise examples of process areas. As we have mentioned in Section 6.2, these detailed content-wise examples also highlighted differences between the approaches. This enabled the participants to have a more objective understanding of different STPI approaches.
- The organization needs to decide whether to select a STPI approach with a model representation (Sections 5.2.1, 5.2.2, 5.2.3) or to use individualized approaches (Section 5.2.4). In our case, the STPI approaches selected had a model representation. Since the model representations (staged vs. continuous) are influenced by CMM/CMMi, we found that there is an element of trust in such approaches. Such approaches are also expected to provide better guidance for assessments.
- If an organization decides to select a STPI approach with a model representation, they then need to decide on a particular model representation, typically staged vs. continuous. As we discuss in Section 8, most organizations prefer a continuous path to improvement as they can easily adapt to the specific needs of the continuous approach.
- An organization needs to know that there are STPI approaches specialized for a specific domain that could be the candidates for selection. However the degree of completion of such approaches need to be assessed beforehand.
- An organization needs to know that for assessment, certain STPI approaches would require an accredited accessor or an experienced external person. This is done to promote transparency and objectivity in assessment results. Also most of the STPI approaches require qualitative data for assessment. This means an assessment of defined processes using interviews, observations and document analysis. It is generally helpful to initially conduct an informal assessment that reflects on the current state of practice in an organization.
- We also realized that for successful selection and application of STPI approaches, extended knowledge in software testing is essential. This could mean different things for an organization, such as having defined roles in software testing, having a test expert or even a dedicated software testing group.

6.6. General information about TPI® NEXT and TMMi®

Here the concepts and especially the specific terminologies of both approaches are introduced to provide better understandability. One significant difference between TMMi and TPI NEXT is their type of model representation. TMMi represents a staged model, whereas TPI NEXT represents a continuous model.

6.6.1. TPI® NEXT

The TPI NEXT model consists of seven elements: key areas, maturity levels, checkpoints, improvement suggestions, enablers, clusters and a test maturity matrix.

6.6.1.1. Key areas. TPI NEXT has 16 key areas. Each key area may have different levels of maturity and the combination of the key areas defines the maturity of the test process as a whole. However, for each key area the maturity is measured individually.

6.6.1.2. Maturity levels. The TPI NEXT model has four maturity levels: initial, controlled, efficient and optimizing. A higher maturity level can only be reached if the preceding maturity level is fulfilled.

6.6.1.3. Checkpoints. Checkpoints are statements regarding the test process. The question whether these stated requirements are satisfied by the investigated test process have to be answered with simple ‘yes’ or ‘no’ replies. A checkpoint always relates to a specific key area and a specific maturity level of the respective key area. A key area is at a certain maturity level when all its checkpoints are satisfied.

6.6.1.4. Clusters. The model enables a stepwise growth from initial to optimizing levels. Each step is indicated by clusters of checkpoints. A cluster is a group of checkpoints from multiple key areas that function as one improvement step. A cluster is used for the purpose of increasing the maturity of the test process. Each cluster is identified by an alphabetic character that represents its position in the improvement path.

6.6.1.5. Enablers. The test process and software development lifecycle model go hand in hand. The enablers help to understand how both processes can benefit from exchanging each other’s best practices.

6.6.1.6. Improvement suggestions. TPI NEXT recommends improvement suggestions and guides an organization to meet checkpoints. The improvement suggestions are practice-based, adaptable and optional to consider.

6.6.1.7. Test maturity matrix. After conducting a test process assessment, the analysis result is shown diagrammatically in a test maturity matrix. This matrix provides an overall picture of the current situation of the test process by highlighting the fulfilled checkpoints of all key areas. Furthermore, the test maturity matrix provides an insight by showing a comparison between its current situation and what level should be achieved in the future to obtain higher maturity.

6.6.2. TMMi®

TMMi consists of five maturity levels: initial, managed, defined, measured and optimization. Each maturity level consists of a set of process areas. The process areas are specific aspects of the testing process that are deemed to be significant for the particular level of maturity. Process areas further consist of three components: Required, expected and informative components.

6.6.2.1. Required components. Required components consist of specific and generic goals which must be achieved to fulfill the requirements of the specific process area and the associated maturity level. Specific goals are specific to a process area whereas generic goals are generally defined statements recurring in all process areas.

6.6.2.2. Expected components. Specific and generic goals are further described by specific and generic practices which belong to the group of expected components. These practices or acceptable alternatives are typically in place to achieve the goals.

6.6.2.3. Informative components. Informative elements can be subpractices, example work products, notes, examples or references. They serve as further information about specific and generic practices.

6.7. Test process assessment using TPI® NEXT and TMMi®

Besides an independent comparison of the content of TPI NEXT and TMMi, the mapping between the two approaches builds a solid basis for the comparison of the results from the application of these approaches. The assessment of the case organization's testing process using TMMi will result in one test maturity level valuing the process as a whole (staged representation). The assessment result of TPI NEXT will be a matrix stating the maturity level for each key area separately (continuous representation).

For the application of TPI NEXT and TMMi the instructions given in v. Ewijk et al. (2013) and van Veenendal (2008) were followed. However, both approaches demand for an assessment done by experienced personal. TPI NEXT either proposes to perform the assessment by an individual who is familiar with the test processes or the BDTPI (Business Driven Test Process Improvement) model or recommends the use of an external expert. In TMMi the requirements in regards to the assessor are even stricter. The TMMi assessment method application requirements (TAMAR) state that a formal assessment can only be performed by an accredited lead assessor. The necessary accreditation can only be gained by the TMMi Foundation.

Assessments without an accredited assessor can only be performed as informal assessments. Formal and informal assessments mainly differ in the presentation of the assessment result. Only formal TMMi assessments allow the statement of the maturity level of the assessed organization. Informal assessments result in a report describing the state of practice in the assessed organization. Due to the absence of an accredited assessor we could base our assessment only on the instructions of an informal assessment. Nevertheless, since the objective of the application of the approaches was to compare their assessment results, we adapted the procedures proposed by the approaches in this direction.

The assessment process of both approaches is generally similar, i.e., collection of data through interviews, data analysis and documentation of results.

The use of the two different approaches in this study was split between two researchers. Researcher A was responsible for the TPI NEXT assessment while Researcher B did the TMMi assessment. However, due to time limits and for the convenience of all participants, we decided to have joint interviews for both approaches.

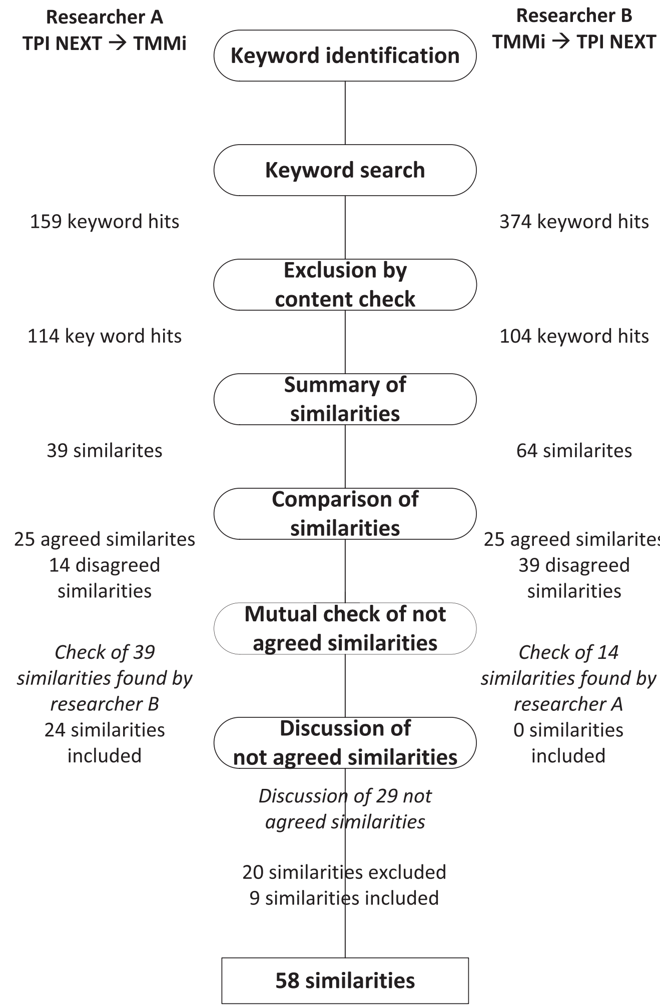

Fig. 6. Mapping between TPI NEXT and TMMi.

7. Case study results

The workshop results given in Section 6.2 resulted in two approaches (TPI® NEXT and TMMi®) that were considered valuable for test process improvements in our case organization. This answers our $RQ_{cs}1$. The following section answers the remaining $RQ_{cs}2$ and $RQ_{cs}3$ of our case study.

7.1. Mapping between TPI® NEXT and TMMi®

In order to compare the results of an assessment, first it is important to compare the approaches to see if they are similar or otherwise. Therefore, a mapping between TPI NEXT and TMMi was done before the actual assessment. The mapping of TPI NEXT and TMMi consisted of checking similarities or differences between the key areas of TPI NEXT and the process areas of TMMi. To obtain triangulation, this mapping was first performed by two researchers individually.

Both researchers followed the same process, but they examined the approaches from different perspectives. Researcher A mapped the content of TPI NEXT to TMMi, while Researcher B mapped the content of TMMi to TPI NEXT. The mapping is illustrated in Fig. 6 and is described as follows:

- Identification of keywords
  Keywords that represent the process areas of TMMi with its specific goals and the key areas TPI NEXT with its checkpoints were identified. Keywords extracted from TMMi level 2 are shown in

Table 30
Keywords extracted from TMMi level 2.

Process area	Specific goal	Keyword
Test policy and strategy		test policy, test strategy
	Establish a test policy	test policy
	Establish a test strategy	test strategy
	Establish test performance indicators	performance indicator, performance, indicator
Test planning		test planning
	Perform a product risk assessment	product risk assessment, risk
	Establish a test approach	test approach
	Establish test estimates	test estimates, estimate, estimating
	Develop a test plan	test plan
	Obtain commitment to the test plan	commitment, test plan
Test monitoring and control		test monitoring, test control, monitoring, control, monitor
	Monitor test progress against plan	progress
	Monitor product quality against plan and expectations	quality
	Manage corrective action to closure	corrective action, closure
Test design and execution		test design, test execution, design, execution
	Perform test analysis and design using test design techniques	test analysis, analysis, test design technique, test design
	Perform test implementation	test implementation, implementation, implement
	Perform test execution	test execution, execution
	Manage test incidents to closure	test incident, incident, closure
Test environment		test environment
	Develop test environment requirements	test environment requirement, test environment, requirement
	Perform test environment implementation	test environment implementation, implementation, implement
	Manage and control test environments	test environment

Table 31
Keywords extracted from TPI NEXT.

Key area	Keywords
Stakeholder commitment	stakeholder, resource, commitment, product risk, test process
Degree of involvement	involvement, involved, lessons learned
Test strategy	test strategy, test level
Test organization	test organization, test policy
Communication	communication, test team
Reporting	report, product risk, lifecycle, test process
Test process management	test plan, evaluation
Estimation and planning	effort, estimation, test plan, dependency, techniques
Metrics	metrics
Defect management	defect, management, monitor, future
Testware management	management, test process, testware, documents
Methodology practice	methodology, test process, test methods, feedback, template
Tester professionalism	tester professionalism, training, test tasks, performance
Test case design	test case, test design, test basis
Test tools	test tool
Test environment	test environment, test environment requirement

Table 30 and the keywords extracted from TPI NEXT are shown in Table 31.

- Search for keywords
  The key words identified in one approach were searched in the other approach. Hits were documented in a matrix that showed the location where the key words were found.
  For better search results, the data basis for the search was extended to specific goals besides process areas in TMMi and checkpoints besides key areas in TPI NEXT. The search of keywords from TPI NEXT in TMMi by Researcher A resulted in 159 hits, and the search of keywords from TMMi in TPI NEXT by Researcher B resulted in 374 hits.
- Exclusion of hits based on their context
  The contents of the process areas (TMMi) and key areas (TPI NEXT) that contained the identical keywords were checked upon whether they convey the same meaning and appear in the same context in both approaches.
  Researcher A excluded 45 keyword hits in which the keywords were not used in the same context in both approaches. Researcher B excluded 270 keyword hits.
- Summary of individually found similarities between TPI NEXT and TMMi
  The extended data basis for the keyword search was now narrowed down to process areas and key areas only. Keyword hits from lower levels were transferred to the corresponding higher levels. The results were summarized to 39 similarities found by Researcher A and 64 similarities found by Researcher B.
- Comparison of individually found similarities
  The mapping results of both researchers were compared. In total, 25 of the found similarities between TPI NEXT and TMMi had been found by both researchers, while 14 similarities had only been found by Researcher A and 39 had only been found by Researcher B.
- Mutual check of not agreed similarities
  All similarities only identified by one researcher were checked by the other researcher. Researcher A checked the 39 similarities that were only identified by Researcher B, and Researcher B checked the 14 similarities that were only identified by Researcher A. In this step Researcher A agreed to include 24 similarities found by Researcher B. Researcher B did not include any similarities in this step.
- Final discussion of not agreed similarities
  The remaining 29 similarities found by only one researcher were now discussed by both researchers. Both researchers presented

their arguments for exclusion or inclusion of these similarities between TPI NEXT and TMMi. In the discussion, the researchers agreed to exclude 20 and to include 9 similarities.

Finally, a total of 58 similarities between TPI NEXT and TMMi were identified. These are presented in Table 32.

For the interpretation of the results it is crucial to take into consideration the different model representations of TPI NEXT and TMMi. TPI NEXT is a continuous approach. Each key area can be assessed individually by all maturity levels. Note that the letters 'C', 'E' and 'O' refer to the three maturity levels of the key areas in TPI NEXT and stand for 'Controlled', 'Efficient' and 'Optimizing'. On the other hand, TMMi is a staged approach. The process areas are linked to the maturity level. Therefore, there are two perspectives in the interpretation of results: (1) TMMi process areas vs. TPI NEXT key areas (2) TMMi maturity levels vs. TPI NEXT maturity levels.

7.1.1. TMMi process areas vs. TPI NEXT key areas

Most of the aspects covered by lower levels of maturity in the key areas of TPI NEXT can be found in the process areas of Maturity Level 2 of TMMi. Exceptions are the key areas 'Testware management', 'Methodology practice', 'Tester professionalism' and 'Test tools'. None of the aspects of these key areas are covered in Maturity Level 2 of TMMi. However, lower maturity aspects of the key areas 'Methodology practice' and 'Tester professionalism' are covered by Maturity Level 3 of TMMi.

The aspects of TPI NEXT's 'Testware management' key area are not covered by TMMi at all. And likewise, the process area 'Quality Control' of TMMi is not addressed by TPI NEXT at all.

7.1.2. TMMi maturity levels vs. TPI NEXT maturity levels

On the contrary, even though aspects of all maturity levels of the TPI NEXT key areas 'Test strategy', 'Test organization', 'Reporting', 'Test process management', 'Estimating and planning', 'Tester professionalism' and 'Test case design' are covered by process areas of TMMi, the maturity levels of these TPI NEXT key areas do not exactly correspond to the respective maturity levels in TMMi. While the aspects of all maturity levels of TPI NEXT's key area 'Test strategy' correspond to TMMi's process areas 'Test policy and strategy' and 'Test planning' in Maturity Level 2 and the aspects of all maturity levels of the key area 'Estimating and planning' in TPI NEXT correspond to 'Test planning' also in Maturity Level 2 of TMMi, the aspects of TPI NEXT's 'Tester professionalism' are reflected by the process areas 'Test organization' and 'Test training program' in Maturity Level 3 of TMMi. Furthermore, the aspects of the key areas 'Test organization', 'Reporting', 'Test process management' and 'Test case design' are corresponding to process areas of different maturity levels of TMMi.

However, most aspects addressed by process areas in higher maturity levels of TMMi (Levels 4 and 5) are accordingly addressed by the highest maturity level (optimizing) in the key areas of TPI NEXT. And likewise, most aspects addressed by process areas in lower maturity levels of TMMi (Levels 2 and 3) are addressed by lower maturity levels (controlled and effective) in the key areas of TPI NEXT.

7.2. Results of test process assessment using TPI NEXT and TMMi

To answer $RQ_{cs}3$, the two approaches TPI NEXT and TMMi were used in parallel to assess the case organization's test process. In particular, we combined the data analysis procedures for TPI NEXT and TMMi presented in Section 6.4 and the mapping between the two approaches presented in Section 7.1.

7.2.1. Elements of test process assessment

Table 33 illustratesthe assessment results of both the TMMi and the TPI NEXT assessment in combination with the mapping results. The fulfillment degree of the process areas in TMMi and the key areas separated by maturity level in TPI NEXT (i.e., C (controlled), E (efficient), O (optimizing)) respectively is indicated by three levels: 'FF' (fully fulfilled), 'PF' (partly fulfilled) and 'NF' (not fulfilled). It is to be noted that for TPI NEXT, in addition to C, E and O, there is another maturity level that is named as 'Initial' but since by default any organization is at this level, we did not consider it in our assessment.

To achieve a rating of 'FF (fully fulfilled)', in TMMi, all specific goals of the respective process area, and in TPI NEXT, all checkpoints of the respective key area, have to be fulfilled. Similarly, if only few of all the specific goals of the respective process area in TMMi are fulfilled or only few of all the checkpoints of the respective key area in TPI NEXT are fulfilled, a rating of 'PF' (partly fulfilled) is achieved. For a rating of 'NF' (not fulfilled), none of the specific goals of the respective process area in TMMi, and for TPI NEXT, none of the checkpoints of the respective key area have to be fulfilled. TMMi process areas that have not been investigated in the case organization are marked with 'NA' (not applicable).

7.2.2. TMMi assessment

The staged model representation of TMMi demands the assessment to begin with the investigation of process areas belonging to Maturity Level 2 named as 'Managed'. Only if all process areas of Level 2 are fulfilled the assessment proceeds with the investigation of process areas belonging to Maturity Level 3. Due to the low level of maturity present in the case organization the assessment was therefore limited to the process areas of Maturity Level 2 only. There are 5 process areas in Maturity Level 2 of TMMi that include 'Test policy and strategy', 'Test planning', 'Test monitoring and control', 'Test design and execution' and 'Test environment'. These process areas are marked with '2' in Table 33 indicating their association to Level 2. The Table 33 also mention rest of the process areas at TMMi Maturity Levels 3, 4 and 5 but as we mentioned before, our assessment was limited to TMMi maturity level 2 only.

The TMMi assessment resulted in all five process areas of Maturity Level 2 being assessed as 'partly fulfilled'. For the first process area 'Test policy and strategy', TMMi specify three specific goals of 'Establish a test policy', 'Establish a test strategy' and 'Establish test performance indicators'. For each of these specific goals, the case organization's test process was assessed with respect to the fulfillment of the respective specific practices recommended by TMMi. All of these specific practices were assessed as being 'partly fulfilled' except specific practice of 'Define test goals' that was assessed as 'not fulfilled', coming under the specific goal of 'Establish a test policy'.

For the second process area of 'Test planning', five specific goals are specified by TMMi, namely 'Perform a product risk assessment', 'Establish a test approach', 'Establish test estimates', 'Develop a test plan' and 'Obtain commitment to the test plan'. All the specific practices relating to each of these specific goals were assessed with respect to fulfillment. All process areas were assessed to be 'partly fulfilled' except 'Obtain commitment to the test plan' that was assessed to be 'not fulfilled'.

The third process area of 'Test monitoring and control' has three specific goals of 'Monitor test progress against plan', 'Monitor product quality against plan' and 'Manage corrective action to closure'. All the specific practices under respective specific goals were assessed as either being 'partly fulfilled' or 'not fulfilled' thus the process area as a whole was assessed as 'partly fulfilled'.

For the fourth process area of 'Test design and execution', there are four specific goals of 'Perform test analysis and design using test design techniques', 'Perform test implementation', 'Perform test execution' and 'Manage test executions to completion'. Same as with the third process area, all the specific practices under respective specific goals were assessed as either being 'partly fulfilled' or 'not fulfilled', thus the fourth process area was assessed as 'partly fulfilled'.

The last process area 'Test environment' has three specific goals, namely 'Develop test environment requirements', 'Perform test

Table 32
Mapping between TPI NEXT and TMMi (‘x’ indicate a similarity).

TMMi Process area		Stakeholder commitment			Degree of involvement			Test strategy			Test organization			Communication			Reporting			Test process management			Estimating and planning			Metrics			Defect management			Testware management			Methodology practice			Tester professionalism			Test case design			Test tools			Test environment		
		C	E	O	C	E	O	C	E	O	C	E	O	C	E	O	C	E	O	C	E	O	C	E	O	C	E	O	C	E	O	C	E	O	C	E	O	C	E	O	C	E	O	C	E	O	C	E	O
2	Test policy and strategy	x						x	x	x		x																																					
2	Test Planning	x				x		x	x								x	x		x	x		x	x	x	x																							
2	Test Monitoring and control		x											x				x		x									x																		x		
2	Test Design and Execution																																								x	x							
2	Test Environment																																														x	x	
3	Test Organization						x				x	x	x									x															x	x	x	x									
3	Test Training Program																																					x											
3	Test Lifecycle and Integration										x								x			x													x														
3	Non-functional Testing																																									x							
3	Peer Reviews																																																
4	Test Measurement																x		x								x																						
4	Product Quality Evaluation		x											x													x																						
4	Advanced Reviews																		x																														
5	Defect Prevention					x																									x											x							
5	Quality Control																																																
5	Test Process Optimization												x						x			x															x						x			x			

x Similarity between TMMi and TPI NEXT

Table 33
Comparison of assessment results done on the mapping between TPI NEXT and TMMi.

TMMi Process area		TPI Next Key area →	Stakeholder commitment			Degree of involvement			Test strategy			Test organization		
			C	E	O	C	E	O	C	E	O	C	E	O
			PF	NF	NF	PF	PF	PF	PF	PF	NF	NF	NF	NF
2	Test policy and strategy	PF	X						X	X	X		X	
2	Test Planning	PF	X				X		X	X				
2	Test Monitoring and control	PF		X										
2	Test Design and Execution	PF												
2	Test Environment	PF												
3	Test Organization	NA						X				X	X	X
3	Test Training Program	NA												
3	Test Lifecycle and Integration	NA										X		
3	Non-functional Testing	NA												
3	Peer Reviews	NA												
4	Test Measurement	NA												
4	Product Quality Evaluation	NA		X										
4	Advanced Reviews	NA												
5	Defect Prevention	NA					X							
5	Quality Control	NA												
5	Test Process Optimization	NA												X

TMMi Process area		TPI Next Key area →	Communication			Reporting			Test process management			Estimating and planning		
			C	E	O	C	E	O	C	E	O	C	E	O
			PF	NF	NF	PF	NF	NF	PF	NF	NF	PF	NF	NF
2	Test policy and strategy	PF												
2	Test Planning	PF				X	X		X	X		X	X	X
2	Test Monitoring and control	PF	X				X		X					
2	Test Design and Execution	PF												
2	Test Environment	PF												
3	Test Organization	NA									X			
3	Test Training Program	NA												
3	Test Lifecycle and Integration	NA						X			X			
3	Non-functional Testing	NA												
3	Peer Reviews	NA												
4	Test Measurement	NA				X		X						
4	Product Quality Evaluation	NA	X											
4	Advanced Reviews	NA						X						
5	Defect Prevention	NA												
5	Quality Control	NA												
5	Test Process Optimization	NA						X			X			

TMMi Process area		TPI Next Key area →	Metrics			Defect management			Testware management			Methodology practice		
			C	E	O	C	E	O	C	E	O	C	E	O
			NF	NF	NF	FF	PF	NF	NF	PF	NF	NF	NF	NF
2	Test policy and strategy	PF												
2	Test Planning	PF	X											
2	Test Monitoring and control	PF				X								
2	Test Design and Execution	PF												
2	Test Environment	PF												
3	Test Organization	NA												X
3	Test Training Program	NA												
3	Test Lifecycle and Integration	NA										X		
3	Non-functional Testing	NA												
3	Peer Reviews	NA												
4	Test Measurement	NA		X										
4	Product Quality Evaluation	NA		X										
4	Advanced Reviews	NA												
5	Defect Prevention	NA						X						
5	Quality Control	NA												
5	Test Process Optimization	NA												X

TMMi Process area		TPI Next Key area →	Tester professionalism			Test case design			Test tools			Test environment		
			C	E	O	C	E	O	C	E	O	C	E	O
			NF	NF	NF	PF	NF	NF	PF	PF	PF	PF	NF	PF
2	Test policy and strategy	PF												
2	Test Planning	PF												
2	Test Monitoring and control	PF										X		
2	Test Design and Execution	PF				X	X							
2	Test Environment	PF										X	X	
3	Test Organization	NA	X	X	X									
3	Test Training Program	NA	X											
3	Test Lifecycle and Integration	NA												
3	Non-functional Testing	NA					X							
3	Peer Reviews	NA												
4	Test Measurement	NA												
4	Product Quality Evaluation	NA												
4	Advanced Reviews	NA												
5	Defect Prevention	NA						X						
5	Quality Control	NA												
5	Test Process Optimization	NA						X			X			

X Similarity between TMMi and TPI NEXT; Similarities in assessment results; Differences in assessment results environment implementation' and 'Manage and control test environments'. None of the specific practices for first and second specific goals were met while for 'Manage and control test environments', few were fulfilled. The overall assessment for the process area was thus 'partly fulfilled'.

7.2.3. TPI NEXT assessment

In contrary to TMMi, the continuous approach of TPI NEXT allows for an assessment of all 16 key areas. Each key area can be at one of the four maturity levels of 'Initial (I)', 'Controlled (C)', 'Efficient (E)' and 'Optimizing (O)'. Due to the difference in model representation (staged vs. continuous), some aspects of the case organization that have been investigated by TPI NEXT and assessed as partly fulfilled ('PF'), have not been investigated by TMMi because they fall beyond TMMi Maturity Level 2. Such TPI NEXT key areas include: 'Degree of involvement', 'Communication', 'Reporting' and 'Test tools'.

In general, the outcome of the TPI NEXT assessment shows a similar result to TMMi assessment. The 16 TPI NEXT key areas were assessed for fulfillment of checkpoints at three maturity levels of 'Controlled', 'Efficient' and 'Optimizing'. As an example, the key area 'Stakeholder commitment' was assessed as 'partly fulfilled' at 'Controlled' level while as 'not fulfilled' at 'Efficient' and 'Optimizing' levels. This is due the case organization not meeting any of the checkpoints for 'Efficient' and 'Optimizing' levels for the key area 'Stakeholder commitment'. One exception in the assessment results for TPI NEXT was the key area of 'Defect management' that was assessed to be 'fully fulfilled' at 'Controlled' level. Rest all the key areas were assessed to be either 'partly fulfilled' or 'not fulfilled' at the three levels of 'Controlled', 'Efficient' and 'Optimizing'. The complete results for all 16 key areas are given in Table 33.

7.2.4. Overlapping concerns

There are some key areas of TPI NEXT in which similarities with TMMi process areas of Level 2 had been identified by the mapping, but which have been assessed as 'not fulfilled' in the TPI NEXT assessment compared to the 'partly fulfilled' rating in TMMi. These are the Efficient level of 'Stakeholder commitment', the Optimizing level of 'Test strategy', the Efficient level of 'Test organization', the Efficient level of 'Reporting', the Efficient level of 'Test process management', the Efficient and Optimizing level of 'Estimating and planning', the Controlled level of 'Metrics', the Efficient level of 'Test case design' and the Efficient level of 'Test environment'.

As mentioned before, the TPI NEXT assessment resulted in one key area being fully fulfilled, namely the Controlled level of 'Defect management'. The mapping between TMMi and TPI NEXT had shown that the process area in TMMi dealing with similar aspects to this key area was 'Test monitoring and control'. Since the process area belongs to Maturity Level 2 it has also been investigated in the TMMi assessment but it was only assessed as 'partly fulfilled'.

For some specific maturity levels of TPI NEXT, key areas that have been assessed as 'partly fulfilled' for the case organization, the mapping between the two approaches had not identified similarities with TMMi process areas. These are the Controlled level of 'Degree of involvement', the Efficient level of 'Defect management', the Efficient level of 'Testware management', the Controlled and Efficient level of 'Test tools', and the Optimizing level of 'Test environment'.

7.2.5. Summary of test process assessment

Below we present a summary of our findings:

- The TMMi assessment at our case organization resulted in all five process areas of Maturity Level 2 being assessed as 'partly fulfilled'. This is shown as 'PF' in second column of Table 33.
- The TPI NEXT assessment at our case organization resulted in all key areas, with an exception of one, to be either 'partly fulfilled' (represented with 'PF' in second row of Table 33) or 'not fulfilled' (represented with 'NF' in second row of Table 33) at the three levels of 'Controlled', 'Efficient' and 'Optimizing' (represented with 'C', 'E' and 'O' in first row of Table 33). The exception was the key area 'Defect management' that was assessed to be 'fully fulfilled' at 'Controlled' level.
- Few key areas in TPI NEXT were assessed as being partly fulfilled ('PF') but were not assessed for TMMi because they belonged to TMMi maturity levels 3 and above. These TPI NEXT key area were: 'Degree of involvement', 'Communication', 'Reporting' and 'Test tools'. These are represented in Table 33 with symbol having diagonal stripes denoting 'Differences in assessment results'.
- Few key areas in TPI NEXT and process areas in TMMi show similarities but at different levels of fulfillment. The following TPI NEXT key areas were assessed as 'not fulfilled', as compared to the 'partly fulfilled' rating in TMMi: the Efficient level of 'Stakeholder commitment', the Optimizing level of 'Test strategy', the Efficient level of 'Test organization', the Efficient level of 'Reporting', the Efficient level of 'Test process management', the Efficient and Optimizing level of 'Estimating and planning', the Controlled level of 'Metrics', the Efficient level of 'Test case design' and the Efficient level of 'Test environment'. These are also represented in Table 33 with symbol having diagonal stripes denoting 'Differences in assessment results'.

8. Discussion

The SLR and the mapping between TMMi and TPI NEXT performed within the case study provide a major general contribution to the body of knowledge with respect to STPI approaches. In this section, we reflect on our findings in this study.

8.1. Discussion on SLR results

Confirmed by experts working in the area, the SLR provided a complete set of approaches. We observed that the research papers about these approaches do not provide sufficient information (see e.g. Swinkels, 2000; Farooq and Dumke, 2008). Majority of the approaches ($\sim$ 61%) do not include assessment instruments (see e.g., MTPF (Karlström et al., 2005), evidence-based (Kasoju et al., 2013), PDCA-based (Xu-Xiang and Wen-Ning, 2010)) which makes the approaches difficult to be applied in industry. $\sim$ 61% of the identified approaches have even only been developed as 'concepts' or 'brief information' (see e.g., MND-TMM (Ryu et al., 2008), MB-VV-MM (Jacobs and Trienekens, 2002)). Another limitation to the general applicability of the approaches is their specialization to a specific domain (see e.g., TPI Auto (TPI, 2004), Emb-TPI (Jung, 2009)). However this specialization to a specific domain is considered important in contexts where existing approaches are lacking, e.g., in the case of embedded software (Jung, 2009). We also found that only few of the newly proposed STPI approaches include case studies, experiments or surveys as a way to validate them, as in Jacobs and Trienekens (2002), Heiskanen et al. (2012), Jung (2009), Taipale and Smolander (2006), Kasurinen et al. (2011) and Karlström et al. (2005).

Based on the origin of the approach and the testing model which builds the framework for the assessment, we divided the approaches into four groups.

The first group consists of TMM and approaches that are based on TMM or that have been influenced by TMM. Since TMM itself has been significantly influenced by CMM, another approach – TIM – has been included in this group that has not explicitly been influenced by TMM rather than by CMM. So one can argue that the approaches in this group are also influenced by CMM.

In contrast, the formation of the second group is less ambiguous. It consists exclusively of TPI and TPI-based approaches. The third group represent standards and approaches related to these standards. The classification within this group was more ambiguous. One approach, the self-assessment framework for ISO/IEC 29119 based on TIM, has been included in this group since the testing model for this approach is provided by the standard ISO/IEC 29119. Viewed from another perspective, this approach could have been also included in the first group since the assessment process is based on TIM. However, the assessment process was not the *primary* criteria of our classification. Finally, the fourth group include all other approaches that do not have a testing model. They present individual assessments which are not built on a predefined framework.

An alternative classification of the approaches could have been done by their model representations which would result in three groups: approaches without a model, approaches with a continuous model representation and approaches with a staged model representation. In such a classification, the individual approaches would have been grouped as approaches without a model while the TPI approaches would have been belonged to the group of approaches with continuous model representation. The remaining approaches, however, would have been a split between continuous or staged model representations. Especially in the TMM-related approaches, both continuous and staged model representations are used. This, in turn, highlights the influence of CMM on these approaches, since CMM provides both a continuous and a staged representation.

One further classification would have been conceivable: qualitative vs. quantitative approaches. But surprisingly, only one approach was identified that used quantitative data for assessment. All the other assessments were done based on qualitative data gained from interviews or surveys. It is evident that the analysis of qualitative data is a preferred assessment technique as it is expected to provide a much more deeper understanding of the phenomenon under study. This tendency to do qualitative analyses is in-line with the statements given by interviewees during the interview phase of this study. It was claimed that the testing process followed is dependent, e.g., on the current situation, the workload or the tester's experience in an area. This individuality of the process makes an unambiguous interpretation of metrics more difficult and therefore the use of qualitative approaches *more* reasonable.

8.2. Discussion on case study results

With respect to the selected STPI approaches to be applied in the case organization, it was clearly reflected that trust in the given methodologies plays an important role in industry. Only few of the approaches identified by the SLR had been known to our industry partner. We found that the best known approaches in industry were TMMi and TPI/TPI NEXT that were eventually selected for the case organization. This finding is in agreement with Heiskanen et al. (2012) where TPI, TPI NEXT and TMMi are given as the most prominent process improvement models for software testing. It could be argued that these are the most commercially promoted ones, therefore the best known in industry. We also agree with Heiskanen et al. (2012) where authors mention TPI NEXT and TMMi to be more managerial in nature rather than emphasizing on technical issues.

Moreover, industry is to a great extent familiar with process improvement frameworks such as CMM/CMMi and demands similar assessments with respect to testing. A formal assessment performed by a lead assessor accredited by the TMMi Foundation provides such an assessment. Therefore, industry trusts in approaches influenced by CMMi. We believe that the awareness of CMM/CMMi in the case organization and the influence of CMMi on TMMi influenced the voting of at least one participant in the static validation step of this study.

It was also an interesting observation that, firstly, approaches based on a testing reference model were selected for application in the case organization and, secondly, approaches with different model representations were selected. We argue that approaches with a model representation provide better guidance for assessments and industry trust their recommended best practices.

The selection of one approach with a continuous model representation (i.e. TPI NEXT) and one with a staged representation (i.e. TMMi) is especially interesting with respect to the performed mapping between the two approaches and comparison of their results. The advantages and disadvantages of these two different representations are often discussed. It is claimed that the continuous approaches, like TPI NEXT, offer more room for improvements in practice (Heiskanen et al., 2012). The ability to focus on individually chosen aspects of the test process provides the freedom to adapt the STPI to the specific needs of the organization; industry seems to realize that as a very valuable characteristic of a STPI approach.

In staged approaches, like TMMi, it seems to be very difficult to fulfill the requirements to achieve the next higher level since all aspects of a maturity level have to be fulfilled as a whole. This is in agreement with previous research done in (Jung, 2009; Oh et al., 2008). Farooq and Dumke (2008) also found that TMM (the predecessor of TMMi) was lacking in adequate guidelines on many process improvement issues when compared with TPI (the predecessor of TPI NEXT). An official survey performed by the TMMi Foundation on the organizations assessed by a formal TMMi assessment states that 11% of the assessed organizations are at initial level and 89% are at Level 2[2]. Therefore, the low TMMi assessment result of the case organization in this study is not surprising. But, on the hand, it might have been expected that the TPI NEXT assessment would have led to a better result. However, due to the results of the mapping between TMMi and TPI NEXT, these similar assessment results are absolutely reasonable.

Despite their different model representations, the mapping between the approaches showed that they principally resemble to a great extent. Apart from smaller differences, they investigate the same aspects of the testing process and they basically categorize specific requirements to the process in the similar level's maturity. On this basis, it is very likely that they come to the same assessment result. A similar conclusion was reached by Kasurinen et al. Kasurinen et al. (2011) where they combined a maturity-level based approach with ISO/IEC 29119 test standard. Using a pilot study, the authors showed that the combination was feasible.

Nevertheless, the mapping and the detailed comparison of the assessment results, indicated that the requirements of the maturity levels in TMMi are much stricter and more difficult to reach than in TPI NEXT. The comparison results showed that some aspects of the testing process covered by lower maturity levels in TPI NEXT and identified as partly fulfilled in the case organization are allocated to much higher maturity levels in TMMi which have not even been investigated due to the non-fulfillment of Maturity Level 2. And furthermore, the mapping showed that some aspects allocated to Maturity Level 2 in TMMi are spread over all three maturity levels of TPI NEXT. Even an achievement of the highest maturity level in TPI NEXT, in regards to these aspects, would still not lead to an achievement of a higher maturity level in TMMi. Moreover, our experience in performing the assessments with both approaches showed that the definitions given for the checkpoints in TPI NEXT are more superficial and provide a lot of freedom for individual interpretations. Whereas, especially, the generic and specific practices, together with the work examples in TMMi give very detailed descriptions of the testing process, which provides a good guidance in conducting the assessment. On the contrary, one can argue that TMMi is more prescriptive and is less flexible to accommodate a variety of test processes that might suit a particular context; in which case TPI NEXT might be a better approach.

However, for the successful application of both approaches, extended knowledge in software testing is essential.

[2] http://www.tmmi.org/pdf/TMMISurvey2012.pdf .

It is worth mentioning that the focus of this paper is on classical STPI approaches. However other software development methodologies have shown to improve software testing. Agile software development in general and extreme programming in particular can improve development quality and productivity, see e.g., Talby et al. (2006) and Ho et al. (2006).

Our case study presents an assessment of the current state of test process at the case organization with respect to the two STPI approaches. However there is no data to show the actual use of these STPI approaches on a short and long term basis. This is an interesting future work that can build on the results of this paper.

9. Threats to validity

Here we use the guidelines from Runeson and Höst (2009) and Robson (2002) to discuss the threats to the validity of our study.

9.1. Construct validity

"This aspect of validity reflect to what extent the operational measures that are studied really represent what the researcher have in mind and what is investigated according to research questions." Runeson and Höst (2009). To make sure that the constructs discussed in the interview questions are interpreted in the same way by the researchers and the interviewees, the transcripts of every interview were verified for correctness by the interviewees.

There is also a threat that assessment results of the two STPI approaches are influenced by individual's personal judgements. Our case study design minimized this threat in several ways. First, although researchers A and B did the assessments individually, the interviews were conducted jointly. Therefore the data analysis was done jointly. As part of data analysis, it was agreed upon when a particular goal or a checkpoint shall be marked as fulfilled. Secondly, the mapping of two approaches before the assessment ensured that the two researchers were interpreting the concepts in a similar way. Thirdly, in case of a conflict between researchers A and B, a third researcher judged on the matter. Lastly, use of data (source) triangulation (interviews, observation and document analysis) helped minimize the threat of individual's interpretation affecting assessment results.

Evaluation apprehension is a social threat about a human's tendency to present herself in a better way when being evaluated. To mitigate this, the interviewees were assured that the data collected in the interviews would be anonymous which helped them to provide honest and realistic information.

9.2. Internal validity

Internal validity refers to the act of establishing a causal relationship between the treatment and the outcome. Two categories of internal validity threats, *maturity* and *selection* (Wohlin et al., 2012), are relevant for our case study. Maturity considers the factors that can affect the reaction of the subject differently (negatively or positively) as time passes. Negative affect being that the subject gets tired or bored during the interview. To mitigate this threat, the duration of the interviews was planned not to exceed 1 h. Selection is the natural variation in human performance and how their behavior is affecting the result. This threat was minimized by asking the 'organization representatives' for help regarding interviewee selection since they were knowledgeable regarding individuals' professional profiles. The second threat was identified while conducting the workshop for static validation where two external participants placed more emphasis on one STPI approach, namely TPI NEXT, due to their experience with it. This threat was mitigated by carefully analyzing other participants' choices and by selecting two approaches in the case organization.

With respect to the SLR, an internal validity threat arises due to unpublished or grey literature which is not made available. To minimize this threat we contacted the authors of the primary studies through email and asked for unpublished literature with respect to STPI approaches.

9.3. External validity

Threats to external validity are conditions that limit the ability to generalize the results of a study to industrial practice. We emphasize that since it is a case study, there is no population from which a statistically representative sample has been drawn (Runeson and Höst, 2009). However we believe our case study results are relevant to other cases having similar characteristics and due to the fact that most of the STPI approaches are domain-independent. Threats to external validity were specifically minimized by selecting different interviewees from different areas (PU and PD), roles and locations (Gothenburg and Bangalore).

With respect to the SLR, we believe our search strategy that consisted of three phases gave us a representative set of primary studies.

9.4. Reliability

Reliability *"is concerned with to what extent the data and the analysis are dependent on the specific researchers."* (Runeson and Höst, 2009). Regarding interviews, they were piloted with two 'organization representatives'. Furthermore, the interview audio recordings were briefly transcribed and the transcripts were sent back to the respective interviewees for confirmation of their correctness. It needs to be mentioned that one cannot be entirely sure about the *interpretations* of such transcriptions. Interpretation is regarded as a complex epistemological concept that is hard to confirm. However we believe that piloting of interviews and authors' experience in the research area (and related terminologies) helped putting the transcriptions in correct context. The interviewees in our case study belonged to different areas within the organization, had different roles, and were located in different countries. The sample therefore had heterogeneity but it cannot be argued that this variation affected the results because all of them were concerned with the same activity, i.e., software testing.

Threats to reliability in conducting SLR are mitigated by providing detailed documentation on different steps such as the search strategy and the study selection process.

10. Conclusions

This study was divided into two parts. In the first part, we conducted a systematic literature review (SLR) to identify available software test process improvement (STPI) approaches. A total of 18 approaches have been found. We observed that many of these approaches lack information such as assessment instruments that make them difficult to be applied in industry. Based on the information extracted from the identified primary studies (such as completeness of development, availability of information and assessment instruments, and domain limitations of the approaches) six generally applicable STPI approaches have been identified – TMM, TMMi, TPI, TPI NEXT, Test SPICE and observing practice. These six approaches mainly differ with regards to the use of testing process reference models and their model representations.

In the second part of this study, we conducted a case study in which, first, two approaches to be applied in the case organization were selected, and second, two parallel assessments of the organization's testing process were performed using these approaches. The approaches used in this case study were TMMi and TPI NEXT. A major distinction between these two approaches is their model representation: TMMi has a staged model while TPI NEXT uses a continuous model. Based on an initially performed mapping between TMMi and TPI NEXT, the assessment results were compared. With both approaches the testing process of the case organization was assessed to be at the 'initial' level. Based on the mapping between the approaches and the comparison of their detailed assessment results, we found out that both approaches have similar and different characteristics. Mostly, they cover the same aspects of the testing process and categorize these aspects to similar levels of maturity. However, a closer look shows that the detailed assessment results differ, particularly due to the different model representations of the two approaches. The requirements of the maturity levels in TMMi are much stricter and more difficult to reach than in TPI NEXT. However the detailed descriptions of generic and specific practices together with the work examples in TMMi provide good guidance on improving the testing process.

The generalizable results of the SLR and the mapping between the two STPI approaches provide, on the one hand, a good basis for further research in this area. There is a need to conduct further case studies comparing assessment results to strengthen the findings and to perform similar mappings between further approaches to extend the knowledge. On the other hand, these results essentially support industry in selecting an approach to improve its software testing processes.

Acknowledgment

This work was funded by The Knowledge Foundation (KKS) through the project 20130085: Testing of Critical System Characteristics (TOCSYC) and by the Swedish Innovation Agency (VINNOVA) through grant 2014-03397 (IMPRINT).

Appendix A. Interview questions

Warm-up questions.

- How old are you?
- What is your educational background?
- How many years of working experience do you have within this organization? Within information technology in general?

Overview of work tasks.

- What is your role in the organization?
- Which systems/applications are you working with?
- Could you please give us an overview of your usual work tasks?

Questions specific to testing.

- How is your work related to testing?
- When you think of the testing you are doing, do you follow a specific testing process?
- Do you follow a specific method?
- How are regression tests and retests done?
- Who is involved in the test processes, inside or outside of your team?
- Do you assign testing tasks to specific persons?
- In which activities is the customer involved?
- Could you please define your stakeholders?
- How is the stakeholder involved in the overall project? And at what time?
- How do you plan for your testing, what are the activities involved in planning, like resource management, etc.?
- Do you have a test plan? What does the test plan include, for example test assignments, test scope, roles or responsibilities?
- Who is involved in planning? Is the customer also involved?
- What are the things you consider when you plan your testing?
- Are you monitoring the testing activities?
- Do you analyze the product risks and do you have a test strategy related to the product risks?
- Could you please explain the differences between your test levels?
- How do you design test cases?
- Do you use specific test design techniques?
- What is the relation between requirements and test cases?
- How do you document the test cases? Do you follow any specific template? Please provide us with an example document.
- Do you have any tools to support testing?
- Is everyone at the same level of knowledge about testing tools within your team?
- How do you handle communication about the project progress amongst your team? How is the communication with the customer done?
- How do you report the testing activity? Please provide us with the document.
- Do you have any metrics to estimate or monitor the test process? How do you record them?
- What is the process to proceed when you find a defect?
- Do you have a defect reporting template? Please provide us with a copy.
- Do you think everyone follows the same process and uses the same resources?
- How does the test environment look like? Who is responsible? How is it maintained?
- Since you do not have any specific role as tester, how did you gain knowledge about testing? Do you take/give any training?

Questions about the interview conduct.

- How do you feel about the duration of the interview?
- Was it difficult to answer the questions?
- We used open ended questions. Would you have preferred 'yes' and 'no' questions?

References

Ares Casal, J.M., Dieste Tubio, O., Garcia Vazquez, R., Lopez Fernandez, M., Rodriguez Yanez, S., 1998. Formalising the software evaluation process. In: Proceedings of the XVIII International Conference of the Chilean Society of Computer Science (SCCC'98). IEEE.

Basili, V.R., 1993. The experience factory and its relationship to other improvement paradigms. In: Proceedings of the 4th European Software Engineering Conference on Software Engineering (ESEC'93). Springer-Verlag, pp. 68–83.

Bertolino, A., 2007. Software testing research: achievements, challenges, dreams. In: Proceedings of the Conference on The Future of Software Engineering (FOSE'07).

Burnstein, I., 2003. Practical Software Testing: A Process-oriented Approach. Springer Inc., New York, NY, USA.

Burnstein, I., Suwanassart, T., Carlson, R., 1996. Developing a testing maturity model for software test process evaluation and improvement. In: Proceedings of the 1996 International Test Conference (ITC'96).

Burnstein, I., T. Suwannasart, A.H., Saxena, G., Grom, R., 1999. A testing maturity model for software test process assessment and improvement. Softw. Qual. Prof. 1 (4), 8–21.

Card, D.N., 2004. Research directions in software process improvement. In: Proceedings of the 28th Annual International Computer Software and Applications Conference (COMPSAC'04). IEEE Computer Society.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Meas. 20 (1), 37–46.

Collofello, J.S., Yang, Z., Tvedt, J.D., Merrill, D., Rus, I., 1996. Modeling software testing processes. In: Proceedings of the 1996 IEEE Fifteenth Annual International Phoenix Conference on Computers and Communications (IPCCC'96).

Dangle, K.C., Larsen, P., Shaw, M., Zelkowitz, M.V., 2005. Software process improvement in small organizations: a case study. IEEE Softw. 22 (6), 68–75.

Ericson, T., Subotic, A., Ursing, S., 1997. TIM – a test improvement model. Softw. Test. Verification Reliab. 7 (4), 229–246.

v. Ewijk, A., Linker, B., van Oosterwijk, M., Visser, B., 2013. TPI®NEXT-Business Driven Test Process Improvement. SogetiBooks.

Farooq, A., Dumke, R.R., 2007. Research directions in verification & validation process improvement. SIGSOFT Softw. Eng. Notes 32 (4), 1–4.

Farooq, A., Dumke, R.R., 2008. Developing and applying a consolidated evaluation framework to analyze test process improvement approaches. In: Software Process and Product Measurement. In: Lecture Notes in Computer Science, 4895. Springer Berlin Heidelberg, pp. 114–128.

Farooq, A., Georgieva, K., Dumke, R.R., 2008. A meta-measurement approach for software test processes. In: Proceedings of the 2008 IEEE International Multitopic Conference (INMIC'08).
Fitzgerald, B., O'Kane, T., 1999. A longitudinal study of software process improvement. IEEE Softw. 16 (3), 37–45.
Garcia, C., Dávila, A., Pessoa, M., 2014. Test process models: systematic literature review. In: Mitasiunas, A., Rout, T., OConnor, R., Dorling, A. (Eds.), Software Process Improvement and Capability Determination. In: Communications in Computer and Information Science, 477. Springer International Publishing, pp. 84–93.
Goodman, L.A., 1961. Snowball sampling. Ann. Math. Stat. 32 (1), 148–170.
Gorschek, T., Garre, P., Larsson, S., Wohlin, C., 2006. A model for technology transfer in practice. IEEE Softw. 23 (6), 88–95.
Harrold, M.J., 2000. Testing: a roadmap. In: Proceedings of the Conference on The Future of Software Engineering (ICSE'00). ACM, New York, NY, USA.
Heiskanen, H., Maunumaa, M., Katara, M., 2012. A test process improvement model for automated test generation. In: Product-Focused Software Process Improvement. In: Lecture Notes in Computer Science, 7343. Springer Berlin Heidelberg, pp. 17–31.
Ho, C.-W., Johnson, M.J., Williams, L., Maximilien, E.M., 2006. On agile performance requirements specification and testing. In: Proceedings of the 2006 Agile Conference (AGILE'06).
Homyen, A., 1998. An assessment model to determine test process maturity. Illinois Institute of Technology, Chicago, IL Ph.D. thesis.
Ilyas, F., Malik, R., 2003. Adhering to CMM level 2 in medium size software organizations in Pakistan. In: Proceedings of the 7th International Multi Topic Conference (INMIC'03), pp. 434–439.
Jacobs, J., v. Moll, J., Stokes, T., 2000. The process of test process improvement. XOOTIC Mag. 8 (2), 23–29.
Jacobs, J.C., Trienekens, J.J.M., 2002. Towards a metrics based verification and validation maturity model. In: Proceedings of the 10th International Workshop on Software Technology and Engineering Practice (STEP'02).
Jung, E., 2009. A test process improvement model for embedded software developments. In: Proceedings of the 9th International Conference on Quality Software (QSIC'09).
Karlström, D., Runeson, P., Nordén, S., 2005. A minimal test practice framework for emerging software organizations. Softw. Test. Verification Reliab. 15 (3), 145–166.
Kasoju, A., Petersen, K., Mäntylä, M.V., 2013. Analyzing an automotive testing process with evidence-based software engineering. Inf. Softw. Technol. 55 (7), 1237–1259.
Kasurinen, J., Runeson, P., Riungu, L., Smolander, K., 2011. A self-assessment framework for finding improvement objectives with ISO/IEC 29119 test standard. In: Systems, Software and Service Process Improvement. In: Communications in Computer and Information Science, 172. Springer, Berlin Heidelberg, pp. 25–36.
Kiiskila, J., 1998. Practical aspects on the assessment of a review process. In: Proceedings of the 24th Euromicro Conference, (EUROMICRO'98).
Kitchenham, B., Charters, S., 2007. Guidelines for performing systematic literature reviews in software engineering. Technical Report EBSE 2007-001. Keele University and Durham University.
Koomen, T., 2002. Worldwide survey on test process improvement. Tech. Rep. Report TPI Survey. http://www.sogeti.fi/upload/NO/Vores%20ybelser/1634Dokument/tpi_survey2002.pdf.
Koomen, T., Pol, M., 1999. Test Process Improvement: A Practical Step-by-step Guide to Structured Testing. Addison-Wesley.
Kulkarni, S., 2006. Test process maturity models—yesterday, today and tomorrow. In: Proceedings of the 6th Annual International Software Testing Conference (STC'06).
Leffingwell, D., Widrig, D., 2003. Managing Software Requirements: A Use Case Approach, 2nd Pearson Education.
Meng, C.-X., 2009. A goal-driven measurement model for software testing process. In: Proceedings of the 2009 International Forum on Information Technology and Applications (IFITA'09).
Oh, H., Choi, B., Han, H., Wong, W.E., 2008. Optimizing test process action plans by blending testing maturity model and design of experiments. In: Proceedings of the Eighth International Conference on Quality Software (QSIC'08).
Pino, F.J., GarcÈ¡a, F., Piattini, M., 2008. Software process improvement in small and medium software enterprises: a systematic review. Softw. Qual. J. 16 (2), 237–261.
Saldaña Ramos, J., Sanz-Esteban, A., García-Guzmán, J., Amescua, A., 2012. Design of a competence model for testing teams. IET Softw. 6 (5), 405–415.
Rana, K.K.R., Ahmad, S.S.U., 2005. Bringing maturity to test. Inf. Prof. 2 (2), 30–33.
Rasking, M., 2011. Experiences developing TMMi® as a public model. In: Software Process Improvement and Capability Determination. In: Communications in Computer and Information Science, 155. Springer Berlin Heidelberg, pp. 190–193.
Reid, S., 2012. The new software testing standard. In: Achieving Systems Safety. Springer, London, pp. 237–255. doi:10.1007/978-1-4471-2494-8_17.
Rinkevics, K., Torkar, R., 2013. Equality in cumulative voting: a systematic review with an improvement proposal. Inf. Softw. Technol. 55 (2), 267–287.
Robson, C., 2002. Real World Research: A Resource for Social Scientists and Practitioner–Researchers. Blackwell Publishing.
Runeson, P., Höst, M., 2009. Guidelines for conducting and reporting case study research in software engineering. Empirical Softw. Eng. 14 (2), 131–164.
Ryu, H., Ryu, D.-K., Baik, J., 2008. A strategic test process improvement approach using an ontological description for MND-TMM. In: Proceedings of the Seventh IEEE/ACIS International Conference on Computer and Information Science (ICIS'08).
Schweigert, T., Nehfort, A., Ekssir-Monfared, M., 2014. The feature set of Test SPICE 3.0. In: Barafort, B., OConnor, R.V., Poth, A., Messnarz, R. (Eds.), Systems, Software and Services Process Improvement. In: Communications in Computer and Information Science, 425. Springer, Berlin Heidelberg, pp. 309–316.
Steiner, M., Blaschke, M., Philipp, M., Schweigert, T., 2012. Make test process assessment similar to software process assessment: the test SPICE approach. J. Softw. Evol. Process 24 (5), 471–480.
Sulayman, M., Mendes, E., 2011. An extended systematic review of software process improvement in small and medium web companies. In: Proceedings of the 15th Annual Conference on Evaluation and Assessment in Software Engineering (EASE'11).
Suwannasart, T., 1996. Towards the development of a testing maturity model. Illinois Institute of Technology, Chicago, IL Ph.D. thesis.
Swinkels, R., 2000. A comparison of TMM and other test process improvement models. Technical Report 12-4-1-FP. Frits Philips Institute.
Taipale, O., Smolander, K., 2006. Improving software testing by observing practice. In: Proceedings of the 2006 ACM/IEEE International Symposium on Empirical Software Engineering (ISESE'06). ACM, New York, NY, USA.
Talby, D., Hazzan, O., Dubinsky, Y., Keren, A., 2006. Agile software testing in a large-scale project. IEEE Softw. 23 (4), 30–37.
Tayamanon, T., Suwannasart, T., Wongchingchai, N., Methawachananont, A., 2011. TMM appraisal assistant tool. In: Proceedings of the 21st International Conference on Systems Engineering (ICSEng'11).
Thomas, M., McGarry, F., 1994. Top-down vs. bottom-up process improvement. IEEE Softw. 11 (4), 12–13.
TMMi Foundation, 2010. Test maturity model integration (TMMi), version 1. URL http://www.tmmi.org/pdf/TMMi.Framework.pdf.
TPI Automotive, Version 1.01, 2004. http://www.tpiautomotive.de/Docs/TPI%20automotive%20version%201.01.pdf.
Unterkalmsteiner, M., Gorschek, T., Islam, A.K.M.M., Chow, K.C., Permadi, R.B., Feldt, R., 2012. Evaluation and measurement of software process improvement: a systematic literature review. IEEE Trans. Softw. Eng. 38 (2), 398–424.
Varkoi, T.K., Makinen, T.K., 1998. Case study of CMM and SPICE comparison in software process assessment. In: Proceedings of the International Conference on Engineering and Technology Management (IEMC'98).
van Veenendal, E., 2008. Test Maturity Model integration (TMMi). Technical Report Version 1.1. TMMi Foundation.
Wang, Y., Dorling, A., Wickberg, H., King, G., 1999. Experience in comparative process assessment with multi-process-models. In: Proceedings of the 25th EUROMICRO Conference (EUROMICRO'99).
Wohlin, C., Runeson, P., Höst, M., Ohlsson, M.C., Regnell, B., Wesslén, A., 2012. Experimentation in software engineering. Computer Science. Springer.
Xin-ke, L., Xiao-Hui, Y., 2009. A goal-driven measurement model for software testing process. In: WRI World Congress on Software Engineering (WCSE'09).
Xu-Xiang, L., Wen-Ning, Z., 2010. The PDCA-based software testing improvement framework. In: Proceedings of the 2010 International Conference on Apperceiving Computing and Intelligence Analysis (ICACIA'10).
Yin, R.K., 2003. Case study research: Design and methods, 3rd SAGE Publications.
Zil-e-Huma, Bano, M., Ikram, N., 2012. Software process improvement: a systematic literature review. In: Proceedings of the 15th International Multitopic Conference (INMIC'12).

Wasif Afzal is a postdoctoral research fellow at Mlardalen University, Sweden. He received his PhD in Software Engineering from Blekinge Institute of Technology in 2011. His research interests are within software testing, prediction and estimation in software engineering and application of artificial intelligence techniques to software engineering problems.

Snehal Alone received her MSc degree in Software Engineering at Chalmers and the University of Gothenburg, Sweden. Her research interests include software testing and software test improvement frameworks.

Kerstin Glocksien received her MSc degree in Software Engineering at Chalmers and the University of Gothenburg, Sweden. Her research interests include software testing and software test improvement frameworks.

Richard Torkar is a professor at Chalmers and the University of Gothenburg. His focus is on quantitative research methods in the field of software engineering. He received his PhD in software engineering from Blekinge Institute of Technology, Sweden, in 2006.