Scientific software test management – A research agenda

Lars Karg, Dirk Voelz

SAP Research, CEC Darmstadt
SAP AG
Bleichstraße 8
64283 Darmstadt, Germany
Lars.Karg@sap.com
Dirk.Voelz@sap.com

Abstract: The application of scientifically derived theories on software testing strongly depends on their ability to provide a value for practice. We provide a comprehensive review on the focus of research in combination with the requirements for their application. The major challenge for practice remains the continuous collection of data for successful method application. Major research gaps lie in a lack of integrated concepts supporting test organisation.

1 Introduction

Quality assurance plays an important role during the software development process. As stated by Boehm and Papaccio [BP88, p. 1469], it is one of the most essential processes of software development, besides the development itself. Estimations of the National Institute of Standards & Technology [Na02] indicate that defective software causes yearly damage of more than one billion dollar in the United States of America alone.

In recent time increased prospect is put into the Industrialisation of Software Development [KTV07], leading to the introduction of new techniques, such as model-driven software development, and generally viewing software development from an increasingly economic perspective. Hence, the deliberate reconfiguration of the software value chains [FF99], a stronger value-oriented development approach [BA05] and, consequently scientific software management are addressed.

Researchers have focused on numerous topics within quality assurance and testing during the past decades. Some of their results have found common ground in the development and quality assurance processes of software development organizations, others not. In this phase of change and innovation towards industrialised structures within software development, we see the opportunity for a comprehensive review on research today and in the past in combination with the requirements for their application in practice to reach superior customer value.

As section 2 will show, quality assurance and testing within the software development process profits from well established theoretical models. Most of them are well integrated into practice. But there are limitations of current research and research adoption. These will be analysed upon interviews with test managers. The findings of this research are presented in section 3.

2 Quality Assurance and Testing within the Software Development Process

During the last decades, test management developed a clear understanding for terminologies and general interdependencies of quality and its costs. Within this context, quality can be seen as a multidimensional construct (cf. [Ga84]) as well as the fulfilment of requirements of different stakeholders (cf. [Se03, p.24]). Such stakeholders could be management, development/testing and the customer. For instance, the management will definitely have the requirement of low software testing costs, whereas customers expect bug free software that fulfils their needs.

In order to support test management reaching this goal, ISO 9126 or the v-model for testing provide well established general concepts. To analyse gaps and opportunities for future research, this paper follows a qualitative, exploratory research approach to gather sufficient data on test management requirements from eight major independent development organizations of a single independent software provider. Thereby, in order to identify current gaps and match research with practice a three stepped approach was followed as presented in figure 1.

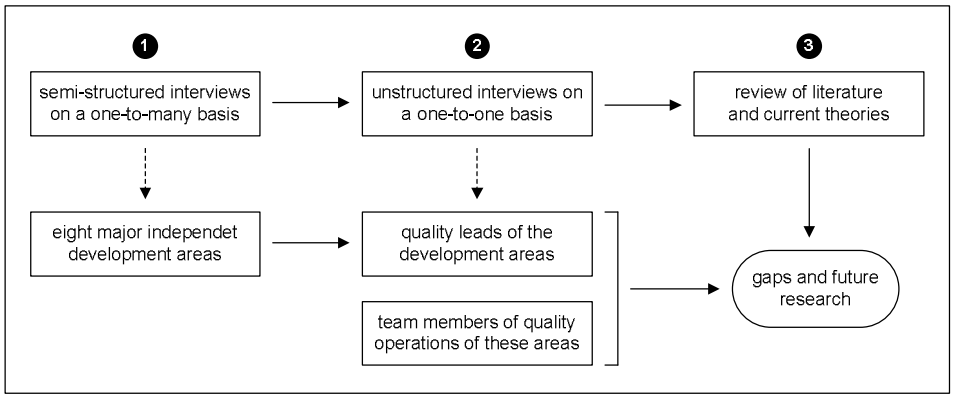

Figure 1: The three stepped research approach.

First all organizations were questioned in semi-structured interviews on a one-to-many base, on what kind of test processes, techniques and test tools they use. This was necessary to get deeper insights into the test organization, but also to develop an objective view on the needs of the organizations and to judge the answers of the quality leads. In the second step, the quality leads of each organization were questioned in unstructured interviews on a one-to-one base upon the status of the application of

scientifically derived methods and tools, their experiences with them, as well as their major requirements for new research. To round their requirements up, two quality specialists from each organization were further asked in unstructured interviews on a one-to-one base about their view on the major requirements of their quality leads. In the final step, the collected requirements were matched with existing theory and analyzed regarding existing gaps. Even though we expect our study to biased by the unique culture of the examined organization, the results provide an application-oriented and promising agenda from future research in test management.

3 Identified areas

For the examined organizations, the application of scientifically founded quality diagnosis models, such as static or dynamic error prediction metrics appear lower than expected. As a reason for this, it is often stated that data gathering itself is complex and the benefit is seldom seen. On the other hand, important aspects of future research concern test techniques and their proper mix in practice as well as integrating models especially considering costs to provide practitioners with use-cases for continuous information collection.

In particular, five major areas were identified, which are seen as promising future research areas. In the following, these areas are named and current gaps are shortly discussed:

1. **Static quality prediction:** In the area of static defect and quality prediction research with sometimes promising results [FP00] have been carried out, but still the outcome are seldom used in practice. A reason is that it is more cost beneficial to depend on developer judgment. Furthermore, research could not solve the threats of reliability [NBZ06] and often "metrics do not capture the essence of whatever it is that they are supposed to measure." [BK04]. As static quality prediction remains an important field for practitioners, metrics capturing this essence are required. Analyzing process indicators might be a first step into this direction.
2. **Dynamic quality prediction:** This area has been addressed by research and is already often applied by practice. Practitioners' requirements can be easily covered if existing theoretical and empirical validated models and techniques are applied properly. Pham [Ph00] for instance, gives an overview of existing techniques for defect and reliability estimations, and their successes. Current challenges for practice lie, besides the right application, in data gathering and not underestimating the benefits of this approach.
3. **Test techniques:** For selected techniques, such as inspections and reviews [La02], regression testing or user-acceptance testing detailed analysis of efficiency and the right application exist [Mu98]. Practice does use the results and techniques quite often. As these techniques appear to be on a quite mature level, less effort is seen for their further improvement, besides of their continuous adoption to new paradigms such as model-driven development.

4. **Integrated models for test mix:** There is the need for integrated models for the optimal test mix, meaning the optimal arrangement of testing techniques within the software development process. From a practitioner's perspective, existing models are very general and, therefore, lack acceptance. Hence, findings are sometimes outdated [Bo81]. Nevertheless, this area is highly relevant. As an example, it is well known that finding and fixing defects in earlier phases save money and increase customer satisfaction. For practitioner's acceptance, however, methods for quantification of these theories for decision support during testing are required.
5. **Integrated cost models:** There are some models dealing with modelling of costs within the software development process, such as CoCoMo II [BH00] and, especially, for quality costs COQUALMO [De99]. But most approaches are too generic or just not applicable. Same conclusion was documented by Wagner [Wa06]. Hence, he proposed a model that covers all analytical quality assurance techniques [Wa06a], but this model again does not fill the gap, for instance, quality management costs are not included (cf. [Ga03, p. 449]), even so, that "poor management can increase costs more rapidly than any other factor" according to Boehm [Bo81, p.486]. In addition to these generic and general approaches, there are some models for selected techniques. Elbaum et al. [EMR02] propose a cost model for regressions tests and Freimut et al. [FBV05] for inspections and reviews, also they are focusing on particular techniques, they are still too generic. Practitioners need models for cost and benefit optimization especially concerning test automation. In contrast, practice has to understand that theory needs data for developing such models. At the moment, practice complains that models do not exist, but also does not see the need for gathering data.

4 Conclusion

The importance of business oriented validation of theoretical concepts is of increasing importance during the Industrialisation of Software Development. The tendency of practitioners to adapt concepts just because they are new is continuously decreasing and instead, the economic value moves into focus. Several areas of research have been identified that have superior potential to find their way to application in daily business. Theory may react focussing on economic perspectives of testing.

Practice is strongly demanding concepts integrating different research streams towards integrated models supporting test management and test organisation. The overall optimization becomes more important than the individual application of a certain technique. In addition, quantitative measures to lead resources and methods in test management are required to support the current experience-driven practice. First measures do exist in particular areas like code complexity. While their use to support daily practice in test management, e.g. the selection of artefacts for code inspections, remains limited due to reliability deficits of the proposed models. The analysis of process metrics appears to be a promising research field. Efforts are required to support test management using statistical measures derived from best-practice supporting tasks that today are only based on experience.

Bibliography

[BA05] Biffl S.; Aurum A.; Boehm B.; Erdogmus H.; Grünbacher P.: Value-Based Software Engineering, Springer, 2005.

[BH00] Boehm, B. W.; Horowitz, E.; Madachy, R.; Reifer, D.; Clark, B. K.; Steece, B.; Brown, A. W.; Chulani, S.; Abts, C.: Software Cost Estimation with CoCoMo II, Prentice Hall, 2000.

[BP88] Boehm, B. W.; Papaccio, P. N.: Understanding and Controlling Software Costs. IEEE Transactions on Software Enginnering, 14(10), 1988: p.1462-1477.

[Bo81] Boehm, B.: Software Engineering Economics, Prentice Hall PTR, 1981.

[De99] Devnani-Chulani, S.: COQUALMO (COnstructive QUALity MOdel): A Software Defect Density Prediction Model. In Project Control for Software Quality, Kusters et al., (Eds.), Shaker Publishing, 2000.

[EMR02] Elbaum, S.; Malishevsky, A. G.; Rothermel, G. (2002). Test Case Prioritization: A Family of Empirical Studies. IEEE Transactions on Software Enginnering, 28(2), 2002.

[FF99] Farbey, B.; Finkelstein, A. (1999). Exploiting supply chain business architecture., http://www.cs.virginia.edu/~sullivan/EDSER-1/PositionPapers/farby.pdf, 1999, last accessed on 2007-01-01.

[Ga84] Garvin, D. A.: What Does "Product Quality" Really Mean?. Sloan Management Review, 26(1), 1984: p.25-43.

[FP00] Fenton, N.E.; Pfleeger, S. L.: Software Metrics: A Rigorous and Practical Approach. PWS Publishing, Boston 2000.

[FBV05] Freimut, B.; Briand, L. C.; Vollei, F.: Determining Inspection Cost-Effectiveness by Combining Project Data and Expert Opinion. IEEE Transactions on Software Engineering, 31(12), 2005: p.1074-1092.

[Ga03] Galin, D.: Software Quality Assurance: From Theory to Implementation. Addison Wesley, 2003.

[KB04] Kaner, C.; Bond, W. P.: Software Engineering Metrics: What Do They Measure and How Do We Know? Metrics 2004.

[KTV07] Kilian-Kehr, R., Terzidis, O.; Voelz, D.: Industrialisation of the Software Sector. To appear in Wirtschaftsinformatik, 2007.

[La02] Laitenberger,O.: A Survey of Software InspectionTechnologies, World Scientific Publishing, 2002: p. 517-555.

[Mu98] Musa, J.D.: Software Reliability Engineering, McGraw-Hill, 1998.

[NBZ06] Nagappan, N.: Ball, T.; Zeller, A.: Mining metrics to predict component failures. ICSE '06. ACM Press, 2006, p. 452-461.

[Na02] National Institute of Standards & Technology – NIST: The economic impacts of inadequte infrastructure for software quality. NIST Planning Report 02-3, 2002.

[Ph00] Pham, H.: Software Reliability. Springer, 2000.

[Se03] Seghezzi, H. D.: Integriertes Qualitätsmanagement - Das St. Galler Konzept. 2nd Edition, Hanser, 2003.

[Wa06] Wagner, S.: A Literature Survey of the Quality Economics of Defect-Detection Techniques. ISESE'06, 2006.

[Wa06a] Wagner, S.: A Model and Sensitivity Analysis of the Quality Economics of Defect-Detection Techniques. ISSTA'06, 2006.