2021 14th IEEE Conference on Software Testing, Verification and Validation (ICST) | 978-1-7281-6836-4/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICST49551.2021.00035

System and Software Testing in Automotive: an Empirical Study on Process Improvement Areas

Fabio Falcini and Giuseppe Lami
Consiglio Nazionale delle Ricerche
Istituto di Scienza e Tecnologie dell'Informazione "A.Faedo"
Pisa, Italy
{fabio.falcini, giuseppe.lami}@isti.cnr.it

Abstract— **In the automotive domain, the development of software-intensive components is mainly demanded to particular suppliers that are required by car manufacturers a.k.a. OEMs (Original Equipment Manufacturer) to measure and, eventually, improve their development process by applying process models such as Automotive SPICE. Automotive SPICE is a widely-applied reference framework providing a set of requested practices in the development life-cycle, including system and software testing.**
This paper aims at contributing in identifying what are the most frequent weaknesses in system and software testing in automotive. The authors present the results of an empirical study aimed at characterizing and analyzing recurrent system and software testing weaknesses in automotive. The authors, as Automotive SPICE assessors, have evaluated requirement management practices on the basis of the evidences gathered from real industrial development projects during a significant number of assessments performed at several organizations world-wide.
This paper is intended to derive a picture of the state-of-the-practice of system and software testing in automotive focusing on the development of software-intensive automotive components. The purpose is to provide researchers and practitioners with a reference for improvement initiatives aimed at solving those weaknesses.

Index Terms—**Automotive, Process Improvement, Automotive SPICE™, System Testing, Software Testing.**

I. INTRODUCTION

Technological innovation deeply changed automobiles in the last years; modern cars can be regarded in all respects more as complex electronically controlled systems than as mechanical/electro-mechanical devices [1] [2] [3]. Electronic systems, more and more complex and connected by CANs (Control Area Networks), control today the main automobile's functionalities [4]. Consequently, software (with increased demand in terms of size and complexity) is today a crucial car component since it is part of car's components called Electronic Control Units (ECU) that control electronically a large number of the vehicle functions (navigation and infotainment included) [5].

The electronics pervasiveness influenced the automobile's design and development paradigms. In particular, the development of software is mainly demanded to ECU and software suppliers (OEMs are lately involved more closely) that range from small-medium organizations to large and structured ones. In this context project management and software engineering, initially underestimated sides of the ECU development projects, have at present taken the attention of whole automotive industry that require projects to meet increasingly demanding timing and quality objectives. In particular, the market expectations (it is fact that the bulk of car issues currently come from electronics and software issues) and technology advances have produced a real need for improvements at managerial and technical levels in order to keep software developments on track, especially for small and medium-sized enterprises (SME).

Automotive SPICE (*Software Process Improvement and Capability dEtermination*) [6] is a model for software process assessment and improvement that is widely used in automotive. Automotive SPICE provides a mean to assess the capability of the ECU suppliers to release products developed by following a technically sound and disciplined process. Automotive SPICE is extensively applied in automotive [2] mainly as a means for qualifying software suppliers by several OEMs [7].

Every year hundreds of SPICE Automotive assessments are carried out worldwide. The results of these assessments represent a valuable source of information on the state of the art in the development of electronics and software in the automotive sector. Despite this potential wide availability of information, in literature there is, to our understanding, scarceness of studies addressing common trends in such a technologically ever-increasing application domain. The reasons may be different, as the confidentiality of data from assessments and the difficulty of collecting data because the existence of many companies involved in the production of software-intensive automotive components and many assessors.

This paper presents an empirical study based on the information gathered by the authors, as qualified Automotive SPICE assessors, during several assessments carried out in the last years worldwide. The study aims at providing a contribution in answering the following questions:

Q.1 What are the most frequently weak system and software testing practices in automotive?

Q.2 Is there any significant difference, in terms of quality of performance, between the system testing and software testing in automotive?

This study relies on full sets of data taken from a sample of 13 Automotive SPICE assessments performed by the authors. The average number of processes assessed in these assessments

978-1-7281-6836-4/21/$31.00 ©2021 IEEE
DOI 10.1109/ICST49551.2021.00035

is 15, and for each assessment several projects may be used as sources of information for determining the Capability Level of each process. For these reasons, the amount of information and process indicators available from the study sample, although not statistically representative, is significant.

The data sample includes information on several processes ranging from the technical ones (belonging to the Engineering category), as for instance Requirements Analysis, Software Design, Software and System Testing, System Integration processes, to the managerial ones as for instance Problem Resolution Management, Change Management, Risk Management. In this paper, we focus on those processes directly addressing system and software testing. According to the Automotive SPICE process reference model, there are five processes directly dealing with system and software testing: System Integration and Integration Test, System Qualification Test, Software Unit Verification, Software Integration and Integration Test, and Software Qualification Test.

Another significant characteristic that enforces the originality and the validity of this empirical study, is the fact that it uses real data from real software development projects collected in the last 3 years. In literature, empirical studies addressing the same topics very often rely on data taken from questionnaires and/or literature review instead of data and indicators from projects [8], [9], [10].

This paper is structured as follows: in Section II the Automotive SPICE model for software process assessment and improvement is presented and its principal components are described. In Section III the methodological approach set up and followed for conducting this empirical study is presented. In Section IV the data, related to the five processes in the scope of such a study, are presented with the support of tables and graphs for understandability and readability purposes. In Section V the results of the study are discussed and in section VI some conclusions are derived and presented.

II. INTRODUCTION TO AUTOMOTIVE SPICE

Automotive SPICE [1] provides a process framework that disciplines, at a high level of abstraction, the software development activities and allows their capability assessment in matching pre-defined sets of numerous process requirements. Automotive SPICE, as a *de-facto* process standard, is used by car manufacturers to push software process improvement among suppliers of software-intensive systems [11]. The purpose of the standard is to provide both a scheme for evaluating the capability of processes involved in the development of software-intensive vehicle components and a path for their improvement. Process capability is defined as a characterization of the ability of a process to meet current or projected business goals. Many OEMs are using this standard to qualify suppliers by requiring to them the achievement of specific ratings [7]. The Automotive SPICE standard provides both the definition of the relevant processes and a mechanism to assign a rating to processes to measure their Capability [6].
Applying Automotive SPICE means first to identify an assessment scope (i.e. a set of the processes taken among the Automotive SPICE Processes set, along with a target rating for each of them), then to collect evidences of the way these processes are actually deployed in projects, and finally, using the specially defined mechanism, to derive the rating in terms of Capability of each of them.

Although each Automotive SPICE assessment can define its own assessment scope, in practice, the reference Automotive SPICE process scope is the one identified by the VDA (Verband der Automobilindustrie e.V.) [12]. It is composed of a subset of the processes in the Automotive SPICE Process Reference Model, each of them with expected Capability Level 2 or more. The VDA Scope is the Automotive SPICE benchmark in automotive and the reference scope used by automotive OEMs for the qualification of suppliers of software-intensive car components as well. In Table 1 the Automotive SPICE Process Reference Model is provided, the processes in bold are those belonging to the VDA scope.

TABLE I. AUTOMOTIVE SPICE PROCESS REFERENCE MODEL AND VDA SCOPE

Process Id. and Name		Process Id. and Name	
ACQ.3	Contract agreement	**SUP.8**	**Configuration Management**
ACQ.4	**Supplier monitoring**	**SUP.9**	**Problem resolution management**
ACQ.11	Technical requirements	**SUP.10**	**Change request management**
ACQ.12	Legal & Administrative Requirements	PIM.3	Process improvement
ACQ.13	Project requirements	SYS.1	Requirement elicitation
ACQ.14	Request for proposals	**SYS.2**	**System requirements analysis**
ACQ.15	Supplier qualification	**SYS.3**	**System architectural design**
MAN.3	**Project management**	**SYS.4**	**System integration and integration test**
MAN.5	Risk management	**SYS.5**	**System qualification test**
MAN.6	Measurement	**SWE.1**	**Software requirements analysis**
SPL.1	Supplier tendering	**SWE.2**	**Software architectural design**
SPL.2	Product Release	**SWE.3**	**Software detailed design and unit construction**
SUP.1	**Quality Assurance**	**SWE.4**	**Software unit verification**
SUP.2	Verification	**SWE.5**	**Software integration and integration test**
SUP.4	Joint Review	**SWE.6**	**Software qualification testing**
SUP.7	Documentation	REU.2	Reuse program management

TABLE I. shows that processes in Automotive SPICE are conveniently grouped and large in number. The rational behind the VDA scope is to limit the impact on the practitioners by selecting the core of the system and software engineering processes and only few additional fundamental managerial processes.

According to Automotive SPICE every process in the assessment scope can be rated according to a scale composed of 6 Levels (ranging from 0 to 5). Level 0 means that the deployment of the process doesn't achieve the expected outcomes and then it is deployed in an incomplete way.

The achievement of Level 1 means that there is evidence that the expected outcomes of that process have been achieved and then the process purpose is achieved as well (Level 1 is said as related to the process Performance). The Levels from 2 to 5 aren't specifically related to the achievement of the process purpose, they are instead related to the measurement of the level of management, control, measurement and improvement of the process practices (Levels 2-5 are related to the process Capability).

III. The Methodological Approach

During the last two decades the authors, in the capacity of qualified Automotive SPICE Principal Assessor (according to the IntACS international assessor certification scheme) [13], have performed dozens Automotive SPICE assessments of organizations producing software-intensive systems for the automotive industry.

Typically, these Automotive SPICE assessments have targeted the VDA scope (or variants of VDA scope) in several domains (e.g. body electronics, lighting, closures, ADAS, ...).

The assessment used in this study are limited to those performed according to the Automotive SPICE ver. 3.1 that is the reference version since 2017. In Annex A the table summarizing, in anonymous way, the sample of assessments used in this study is provided. Although the sample is limited in number (13) and geographical distribution (Europe, Asia, and North America) it can be considered meaningful by all means because the data set is composed of systematically recorded data from real assessments of real software development projects. Yet the following outcomes have not a statistical validity and are based on empirical observations. In Annex A, the column "Company Size" of has been left void for confidentiality reasons (the indication of company size could lead to the identification of the company itself).

The available data target in total 23 projects (some of them having to comply with ISO 26262 requirements [14]). From a size point of view the organizations ranges from small, medium and large ones.

During Automotive SPICE assessments, evidences and data on the processes in scope are gathered by various means, including interviews, project documents and work products analysis and these data are used to assess (using the expert judgment of the assessors as well) a set of indicators provided by the Automotive SPICE model itself. The assessment results and ratings shall be justified by the collected evidences. These indicators are the so-called Base Practices (process-specific indicators) and the Generic Practices (indicators that are applicable to all processes). Base Practices are indicators of the performance of a specific process, i.e. they represent the set of practices necessary to fulfill the purpose of the process they refer to (Capability Level 1). Generic Practices (GPs) are indicators of the capability of a process that are referred to the level of management, control, measurements, and continuous improvement (Capability Levels 2-5). GPs are out of the scope of this study. In the context of process improvement, it is important to remark that the assessment activity is not limited to a mere rating of process indicators, but it includes also the provision of high-level improvement guidance for the projects under assessment. Assessments represent opportunities for the assessors as they expose them to precious "behind-doors" experience of real projects.

This study is based on the data taken from Automotive SPICE Assessments performed in the time interval 2018 – 2020 using Automotive SPICE version 3.1.

The following step-wise approach is adopted in this study:

S.1 the rating achieved by the Base Practices of the processes under investigation in all the Automotive SPICE assessments considered in this study (Annex A) have been reported in tabular format.

S.2 The software and system testing-related practices having higher frequency of unsatisfactory ratings (i.e. corresponding to BPs achieving a lower rating according to the Automotive SPICE Measurement Framework) have been identified with the support of statistical techniques.

S.3 The rationales of Base Practices weaknesses have been investigated and analyzed in order to identify possible significant trends and commonalities in software and system testing in automotive.

Confidentiality issues has been considered and carefully addressed. The NDAs used for Automotive SPICE assessments usually do not contain explicit non-disclosure clauses about the ratings of the process indicators and, as such, the authors entirely assume the responsibility for providing, albeit anonymously, the process attribute ratings in this paper.

IV. Study Outcomes

As stated above, this paper focuses on system and software testing-related practices. Automotive SPICE has been conceived to include five processes directly addressing testing, two of them address system testing (SYS.4 System Integration and Integration Test and SYS.5 System Qualification Test), the others software testing (SWE.4 Software Unit Verification, SWE.5 Software Integration and Integration Test, and SWE.6 Software Qualification Test). According to [6] the purpose of the SYS.4 process is *"to integrate the system items to produce an integrated system consistent with the system architectural design and to ensure that the system items are tested to provide evidence for compliance of the integrated system items with the system architectural design, including the interfaces between system items"*. The purpose of the SYS.5 process is *"to ensure that the integrated system is tested to provide evidence for compliance with the system requirements and that the system is ready for delivery"*. The purpose of the SWE.4 process is *"to verify software units to provide evidence for compliance of the software units with the software detailed design and with the non-functional software requirements"*. The purpose of the SWE.5 process is *"to integrate the software units into larger software items up to a complete integrated software consistent with the software architectural design and to ensure that the software items are tested to provide evidence for compliance of the integrated software items with the software architectural design, including the interfaces between the software units and* *between the software items"*. Finally, the purpose of SWE.6 is "*to ensure that the integrated software is tested to provide evidence for compliance with the software requirements"*.

This study addresses these five processes in order to provide an answer to the questions stated in Section 1. This section is structured in three sub-sections: the first is aimed at describing the Base Practices of interest for this study, the second aims at presenting the raw resulting data of the study, and the third is devoted to the presentation of study outcomes aggregations.

A. System and Software Testing Practices in Automotive SPICE

Automotive SPICE includes processes dealing with system and software testing. It requires explicitly the existence of several testing phases scoping both software and system verification. The reason for that is that Automotive SPICE has been conceived for targeting suppliers of ECUs (Electronic Control Units) for cars. ECUs are composed of a hardware part with embedded software [1], [15]. In such a context, the SYS.4 and SYS.5 are related to the verification of the functional and non-functional characteristics of the ECU, SWE.4, SWE.5, SWE.6 are related to the software verification only.

In the following Tables the Base Practices of the five processes are listed and shortly described.

TABLE II. SOFTWARE UNIT VERIFICATION PROCESS BASE PRACTICES [6]

SWE.4 Software Unit Verification Process Base Practices	
Id.	Definition
BP1	Develop software unit verification strategy including regression strategy.
BP2	Develop criteria for unit verification in a unit test specification.
BP4	Test software units using the unit test specification according to the software unit verification strategy
BP5	Establish bidirectional traceability between the software detailed design and the unit test specification. Establish bidirectional traceability between the unit test specification and unit test results
BP6	Ensure consistency between the software detailed design and the unit test specification
BP7	Summarize the unit test results and communicate them to all affected parties

TABLE III. SOFTWARE INTEGRATION & INTEGRATION TESTS PROCESS BASE PRACTICES [6]

SWE.5 Software integration & Integration Tests process Base Practices	
Id.	Definition
BP2	Develop a strategy for testing the integrated software items following the integration strategy including regression tests strategy.
BP3	Develop the test specification for software integration test including the test cases according to the software integration test strategy for each integrated software item.
BP5	Select test cases from the software integration test specification assuring sufficient coverage according to the software integration test strategy and the release plan
BP6	Perform the software integration test using the selected test cases. Record the integration test results and logs
BP7	Establish bidirectional traceability between elements of the software architectural design and software integration test cases.
BP8	Ensure consistency between elements of the software architectural design and software integration test cases
BP9	Summarize the software integration test results and communicate them to all affected parties

TABLE IV. SOFTWARE QUALIFICATION TEST PROCESS BASE PRACTICES [6]

SWE.6 Software Qualification Test Process Base Practices	
Id.	Definition
BP1	Develop software qualification test strategy including regression test strategy consistent with the project plan and the release plan.
BP2	Develop the specification for software qualification test including test cases based on the verification criteria, according to the software test strategy to provide evidence for compliance of the integrated software with the software requirements
BP3	Select test cases from the software test specification. The selection of test cases shall have sufficient coverage according to the software test strategy and the release plan.
BP4	Test the integrated software using the selected test cases. Record the software test results and logs
BP5	Establish bidirectional traceability between software requirements and software qualification test cases
BP6	Ensure consistency between software requirements and software qualification test cases
BP7	Summarize the software qualification test results and communicate them to all affected parties

TABLE V. SYSTEM INTEGRATION & INTEGRATION TESTS PROCESS BASE PRACTICES [6]

SYS.4 Software integration & Integration Tests process Base Practices	
Id.	Definition
BP2	Develop a strategy for testing the integrated system items following the integration strategy including regression tests strategy.
BP3	Develop the test specification for system integration test including the test cases according to the system integration test strategy for each integrated system item.
BP5	Select test cases from the system integration test specification assuring sufficient coverage according to the system integration test strategy and the release plan .
BP6	Perform the system integration test using the selected test cases. Record the integration test results and logs
BP7	Establish bidirectional traceability between elements of the system architectural design and system integration test cases
BP8	Ensure consistency between elements of the system architectural design and system integration test cases
BP9	Summarize the system integration test results and communicate them to all affected parties

TABLE VI. SYSTEM QUALIFICATION TEST PROCESS BASE PRACTICES [6]

SWE.6 Software Qualification Test process Base Practices	
Id.	Definition
BP1	Develop system qualification test strategy including regression test strategy consistent with the project plan and the release plan.
BP2	Develop the specification for system qualification test including test cases based on the verification criteria, according to the system test strategy to provide evidence for compliance of the integrated system with the system requirements
BP3	Select test cases from the system test specification. The selection of test cases shall have sufficient coverage according to the system qualification test strategy and the release plan.
BP4	Test the integrated system using the selected test cases. Record the system qualification test results and logs
BP5	Establish bidirectional traceability between system requirements and system qualification test cases
BP7	Ensure consistency between system requirements and system qualification test cases
BP8	Summarize the system qualification test results and communicate them to all affected parties

B. Study Data Report

TABLE VII. reports, for each assessment, the ratings assigned to the Base Practices. Columns correspond to the assessments in the study sample. To be noticed that there is no correspondence between the numeration of the assessments in TABLE VII. columns, and the order used in Annex A to list the Organizational Units involved in the study sample. In this way, we assure confidentiality as it is impossible to associate process indicator ratings to a specific assessment. Each row of the table corresponds to one Base Practice associated with a testing-related process. The cells of the table contain the corresponding rating given to a Base Practice in a specific assessment. The rating values reported in TABLE VII are calculated according to the measurement scale provided by Automotive SPICE to rate the process indicators. Such a rating scale is composed of 4 values (N, P, L, F). Automotive SPICE provides a process indicators rating mechanism that associates a value in the four-value N-P-L-F scale to a percentage of achievement of the Base Practice. In practice, the assessor shall gather enough evidences to establish at what extent a Base Practice is performed, this extent is required to be expressed in percentage. The establishment of an exact percentage of the performance of a practice is very hard to achieve, as it is not about a quantitative measure but it is essentially about a professional judgment based on evidences. Therefore, Automotive SPICE provides, in order to make assessment rating more repeatable and comparable, a mapping between percentages and rating values on the N-P-L-F scale shown in TABLE VIII.

TABLE VIII. AUTOMOTIVE SPICE RATING CORRESPONDENCE

Performance Percentage Range	0%-15%	16%-50%	51%-85%	86%-100%
Rating value	N	P	L	F

Then, if the percentage of performance of a certain Base Practice is evaluated, for example as 70%, the rating to be assigned to that Base Practice is L, if the percentage of performance is evaluated as 25% the rating is P, and so on.

TABLE VII. SYSTEM AND SOFTWARE TESTING-RELATED BASE PRACTICES RATINGS

Base Practices	**Assessment**												
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
SYS.4.BP2	F	L	L	F	-	P	F	-	-	F	F	-	L
SYS.4.BP3	F	L	L	L	-	L	F	-	-	F	F	-	L
SYS.4.BP5	F	L	L	L	-	L	F	-	-	F	F	-	L
SYS.4.BP6	F	L	L	F	-	P	F	-	-	F	F	-	L
SYS.4.BP7	F	L	L	F	-	L	F	-	-	F	F	-	F
SYS.4.BP8	L	L	L	F	-	L	F	-	-	F	F	-	L
SYS.4.BP9	F	L	F	F	-	L	F	-	-	F	F	-	L
SYS.5.BP1	F	F	L	F	F	P	F	-	-	F	F	-	F
SYS.5.BP2	F	F	L	F	F	L	F	-	-	F	F	-	F
SYS.5.BP3	F	F	L	F	F	L	F	-	-	F	F	-	F
SYS.5.BP4	F	L	L	L	F	P	F	-	-	F	F	-	F
SYS.5.BP5	F	F	L	F	F	L	F	-	-	F	F	-	F
SYS.5.BP7	F	L	L	F	F	L	F	-	-	F	L	-	F
SYS.5.BP8	F	F	F	F	F	L	F	-	-	P	F	-	F
SWE.4.BP1	L	P	P	L	L	P	F	F	L	F	F	F	L
SWE.4.BP2	L	P	P	L	F	L	F	F	F	F	F	F	F
SWE.4.BP4	L	P	P	L	L	P	F	F	L	F	F	F	L
SWE.4.BP5	L	L	L	F	F	L	F	F	L	F	L	L	L
SWE.4.BP6	L	L	L	F	F	L	L	F	L	F	L	L	L
SWE.4.BP7	L	L	L	L	L	L	F	F	F	F	F	F	F
SWE.5.BP2	L	P	F	F	L	P	F	F	L	F	F	F	L
SWE.5.BP3	L	P	P	P	L	L	F	L	F	F	F	F	L
SWE.5.BP5	L	P	L	L	F	L	F	L	F	F	F	F	L
SWE.5.BP6	L	P	F	F	L	P	F	F	L	F	F	L	L
SWE.5.BP7	L	L	L	F	L	L	F	L	L	L	F	F	L
SWE.5.BP8	L	L	L	F	L	L	F	L	L	L	F	F	L
SWE.5.BP9	L	L	F	F	L	L	F	F	F	F	F	L	F
SWE.6.BP1	F	P	L	L	F	L	F	F	F	F	F	F	F
SWE.6.BP2	F	L	L	F	L	F	F	F	F	F	F	F	F
SWE.6.BP3	F	L	F	F	L	F	F	F	F	F	F	F	F
SWE.6.BP4	F	P	L	L	F	L	F	F	F	F	F	L	L
SWE.6.BP5	F	L	F	F	F	F	F	F	L	F	F	F	F
SWE.6.BP7	F	L	F	F	L	F	F	F	L	F	F	L	L
SWE.6.BP8	F	L	F	F	F	L	F	F	F	F	F	L	F

To be noticed that, as shown in TABLE VII., the available data for processes related to System Testing are less than those for processes related to Software Testing. This is due to the variability of assessment scopes in the study sample.

In order to facilitate the analysis of the data the rating value of each Base Practice, originally expressed by a value in the four-value scale N-P-L-F, is substituted by a numeric value. To do that, we introduce the following assumption: we consider the mean value of each percentage range and we substitute it to the correspondent N-P-L-F value. According to this mechanism, the N rating will be substituted with the value 0,075 (7,5%), P with 0,33 (33%), L with 0,66 (66%), and F with 0,925 (92,5%).

According to this assumption, it is possible to calculate the average value of the ratings of each Base Practice in the sample of this study. The average values are represented in graphical format in Figure 2.

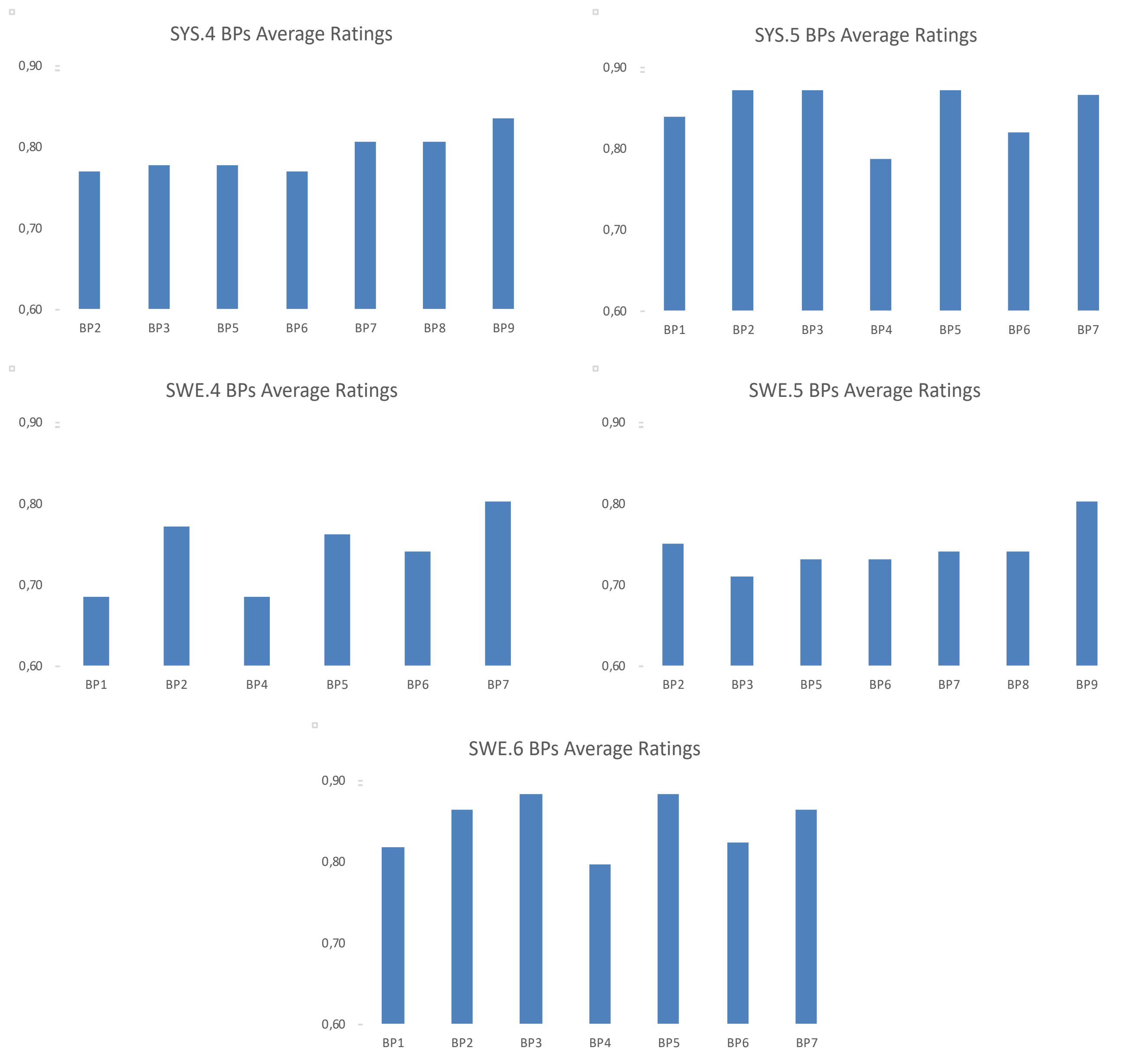

Figure 2: Average Rating of Single Base Practices

C. Study Data Aggregation

A characteristic of the set of processes related to System and Software Testing in Automotive SPICE is that they contain Base Practices with similarities. In particular, we can notice that the Base Practices of these processes can be grouped into 5 clusters that represent homogeneous areas of practice. In other words, each cluster is composed of practices conceptually addressing homogeneous topics. Clusters represent data aggregations that allow to simplify and harmonize the results of this study.

The clusters that can be identified are:

C1 Define a testing strategy
C2 Identify/Specify test cases
C3 Perform testing
C4 Record, report and communicate test results
C5 Establish traceability and assure consistency.

In TABLE IX the Base Practices belonging to each cluster are reported both for software and system testing.

TABLE IX. CLUSTER COMPOSITION

Cluster	System Testing BPs	Software Testing BPs
C1	SYS.4: BP2 SYS.5: BP1	SWE.4: BP1 SWE.5: BP2 SWE.6: BP1
C2	SYS.4: BP3, BP5 SYS.5: BP2, BP3	SWE.4: BP.2 SWE.5: BP3, BP5 SWE.6: BP2, BP3
C3	SYS.4: BP6 SYS.5: BP4	SWE.4: BP4 SWE.5: BP6 SWE.6: BP4
C4	SYS.4: BP9 SYS.5: BP7	SWE.4: BP7 SWE.5: BP9 SWE.6: BP7
C5	SYS.4: BP7, BP8 SYS.5: BP5, BP6	SWE.4: BP5, BP6 SWE.5: BP7, BP8 SWE.6: BP5, BP6

Interesting outcomes can be derived by aggregating the ratings of Base Practices belonging to the same cluster and then calculating the mean value of the ratings. The results are represented in graphical format in Figure 3, where for each testing-related process, the average rating of the 5 Clusters is identified.

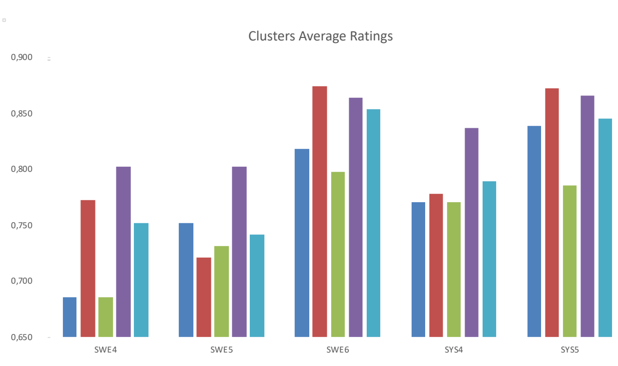

Figure 3. Cluster Average Ratings

The information that can be derived from the sole average calculation can be used to identify the testing phases that can be considered, generally speaking, weaker in practice. The evidences from Figure 3. will be discussed in more details the next section.

The data gathered in the study can be used to further combine the ratings of the clusters of the processes related to software testing processes (SWE.4, SWE.5, and SWE6) and then calculate corresponding average rating. The same for the clusters of the processes related to system testing (SYS.4 and SYS.5). The results are shown in Figure 4. In this way it possible to get an indication of the differences in terms of process capability between the software and system testing practices. Also, the evidences from Figure 4. will be discussed in more details the next section.

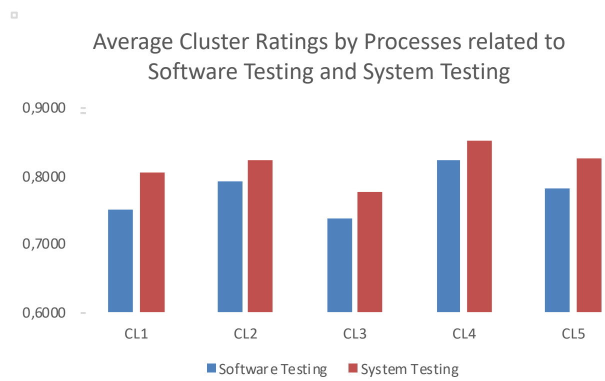

Figure 4. Average Ratings by Base Practices Clusters

V. RESULTS DISCUSSION

This paper presents an empirical study is aimed at identifying and discussing possible recurrent weak and strong areas in the overall testing process in automotive. The study has been carried out using data from assessment performed according to Automotive SPICE ver. 3.1 using real data from real software development project. The study uses data gathered from 13 Automotive SPICE Assessment performed in the last 3 years by the authors.

The overall testing process is addressed by five different processes in Automotive SPICE: System Integration and Integration Test, System Qualification Test, Software Unit Verification, Software integration and Integration Test, and Software Qualification Test processes. It has been noticed that the Base Practices associated with these processes can be clustered by the testing activity they refer to. In total 5 clusters have been identified for each process. Each cluster contains Base Practices that address the same kind of activity even though mapped on different BPs.

The results of the study, reported in detail in this paper, can be summarized to provide an answer to the initial research questions.

A. Research Question Q1 Answers

The outcomes of the study show that the ratings of the Base Practices are not homogeneous. Some Base Practices are weaker than others. In particular, the weaker Base Practices can be found in SWE.4 “Software Unit Verification” process. In this case, SWE.4.BP1 “Develop software unit verification strategy” and SWE.4 BP3 “Test software units” have an average rating lower than 0,70.

This is in line with the experience of the authors, in fact, often software unit verification activities (including static verification of code and unit testing) are performed quite informally, with little available evidences (e.g. test case specification, verification log, verification report, …) and the strategy is often insufficiently defined because it is devoted to the experience and skill of the verifier who works without defined targets and procedures.

Another indication that arises from the collection of the Base Practices average ratings is the general lower average ratings of the SWE.5 “Software Integration and Integration Test” process. This is due to the fact that such a process is sometimes not completely deployed as the testing is principally focused on software units and, even more, on software qualification test (i.e. the testing against functional requirements).

From the analysis of the results of the Base Practices clustering and the related average ratings (summarized in Figure 3.) additional interesting considerations addressing the research question Q1 can be made:

- The Cluster CL5 “Establish Traceability and Assure Consistency” is significantly weaker for SWE.4, SWE.5 processes and also for SYS.4 with respect to the others.
  This can be justified by the fact that the requested traceability for these three processes is between architectural design and integration test cases, while for the other processes the requested traceability is between requirements and test cases. The reason of this weakness, in the authors’ experience, is due to both methodological and technological factors. Creating relationships with architectural elements is not a straightforward procedure as for requirements and test. In fact, architectural descriptions include a variety of approaches including graphical ones. In addition, while it is a common practice to use requirements management supporting tool for deploying traceability between requirements and test cases, in the case of traceability between architectural design elements and test cases there is a limited use of equivalent supporting tools.
- Similarly, the Cluster CL1 “Define a Testing Strategy” is significantly weaker for SWE.4, SWE.5 processes and also for SYS.4 with respect to the others.
  Defining a testing strategy means to decide how to perform testing and how to provide evidence for compliance of the software/system under test. Differently from system/software qualification testing in which a testing strategy is very often clearly stated (often because contractual constraints), in the case of software unit testing and system/software integration testing the related activities are performed sometimes without an explicit definition of the test scope, the methods for test case and test data definition, the criteria to select test cases to run, the characteristics of the testing environment, the eventual coverage target, and the regression testing policy. Such a level of informality is, in the authors’ experience, the main reason for this weakness.
- The Cluster 3 “Perform Testing” is significantly weak for the SWE.4 process. Software unit testing is performed at times quite informally (i.e. without a full specification of test cases and a complete record of test results). This is the main reason for such a weakness. Moreover, in some circumstances, the defined test plan is not completely applied due to timing delays with respect to the project scheduling. Similar considerations are valid also for the Cluster 3 for SWE.5 process, that is weaker as well.

B. Research Question Q2 Answers

The analysis of the results of the average ratings of clusters considering separately the whole software testing area and the and system testing one, provides interesting indications that can be used to answer research question Q2. Data represented in Figure 4. show that, in general, the average clusters rating is higher for the combined system testing related processes than the software testing ones.

This indication shows that the system testing is more mature with respect software testing. Often system testing and software testing are performed by separate teams using different tools and applying different procedures. A possible reason for this evidence is the fact that system engineering is a traditional discipline in automotive with a more consolidated technical and technological background with respect software testing. A second one is linked to economic factors: it is intentional to apply limited focus on software testing due to limited project resources.

VI. Conclusions

The study presented in this paper does not claim to rely on a statistical valid set of data and consequently it doesn’t claim a statistical validity of the results. The study relies on a data taken from Automotive SPICE process assessments performed by the authors on a sample of 13 companies worldwide in the last seven years. These data include evidences collected during the assessments related to procedures, work products, tools, software product characteristics, quality and management indicators. The data available for system testing are less than those for software testing. This is due to the fact that the scope of the assessments is not homogeneous and depends on the business of the organization unit assessed.

The principal originality of this study is the use of real data from real software development projects in automotive. In literature, the empirical studies addressing similar topics are mainly based on literature reviews and surveys typically made by means of questionnaires.

The results of this study can represent a contribution in the identification of the most critical practices in automotive software development projects and can represent both a benchmark for automotive software players and a starting point for setting up process improvement initiatives.

The authors' aim is to continue this study by extending the analysis to other processes available in the data sample and investigating possible correlations among Base Practices. The sample will be used also to find out possible characterizations of the weaknesses in terms of company size, the geographical location and the specific product domain.

REFERENCES

[1] Mössinger, J. (2010). Software in automotive systems. IEEE software, 27(2), 92-94.

[2] Vlaović, J., Vidaković, M., Kovačević, M., Kovačević, B., & Lukić, N. (2020). Developing Consumer Electronics Software Using A-SPICE: Infotainment Development Experience. IEEE Consumer Electronics Magazine, 9(4), 104-110.

[3] Vdovic, H., Babic, J., & Podobnik, V. (2019). Automotive software in connected and autonomous electric vehicles: A review. IEEE Access, 7, 166365-166379.

[4] C. Ebert and J. Favaro, "Automotive Software," in IEEE Software, vol. 34, no. 3, pp. 33-39, May-Jun. 2017, doi: 10.1109/MS.2017.82.

[5] Staron, M. (2017). Automotive software architectures. Automot. Softw. Archit, 33-39.

[6] VDA QMC Working Group 13 / Automotive SIG Automotive SPICE Process Assessment / Reference Model Ver. 3.1, 2017

[7] Fabbrini, F., Fusani, M., Lami, G., Sivera, E.: A SPICE-based Supplier Qualification Mechanism in Automotive Industry. Software Process Improvement and Practice Journal. 2007; 12; 523-528. Wiley InterScience.

[8] Niazi, M., Wilson, D., & Zowghi, D. (2006). Critical success factors for software process improvement implementation: an empirical study. Software Process: Improvement and Practice, 11(2), 193-211.

[9] Niazi, M., Babar, M. A., & Verner, J. M. (2010). Software Process Improvement barriers: A cross-cultural comparison. Information and software technology, 52(11), 1204-1216.

[10] Hall, T., Rainer, A., & Baddoo, N. (2002). Implementing software process improvement: an empirical study. Software Process: Improvement and Practice, 7(1), 3-15.

[11] Mueller, M., Hoermann, K., Dittmann, L., & Zimmer, J. (2008). Automotive SPICE in Practice: surviving implementation and assessment. Rocky Nook, Inc..

[12] VDA: Automotive SPICE Guidelines. 1st Ed. H. Druck, 2017.

[13] INTACS Homepage, www.intacs.info, last accessed 2020/12/30.

[14] ISO 26262 - Road Vehicles - Functional Safety, International Organization for Standardization, 2011.

[15] Denton, T.: Automobile Electrical and Electronic Systems. 5th Ed. Routledge (2004).

Annex A

Study Sample summary:

Domain	Project Team Size	Location	Year	Processes in the Assessment Scope (in bold the process addressed in this paper)
Driving Assistance	50+	Asia	2020	SWE.1, SWE.2, SWE.3, **SWE.4, SWE.5, SWE.6**, MAN.3, MAN.5, SUP.1, SUP.8, SUP.9, SUP.10, SPL.2
HVAC	10	Europe	2020	SYS.1, SY2.2, SYS.3**, SYS.4, SYS.5**, SWE.1, SWE.2, SWE.3, **SWE.4, SWE.5**, **SWE.6**, MAN.3, MAN.5, SUP.1, SUP.4, SUP.8, SUP.9, SUP.10, SPL.2, ACQ.4
body electronics	10	Europe	2019	SY2.2, SYS.3, **SYS.4, SYS.5**, SWE.1, SWE.2, SWE.3, **SWE.4, SWE.5, SWE.6**, MAN.3, SUP.1, SUP.8, SUP.9, SUP.10
body electronics	20	Asia	2018	SYS.1, SY2.2, SYS.3, **SYS.4, SYS.5**, SWE.1, SWE.2, SWE.3, **SWE.4, SWE.5, SWE.6**, MAN.3, MAN.5, SUP.1, SUP.2, SUP.4, SUP.8, SUP.9, SUP.10, SPL.2, ACQ.4
ECU	10	Europe	2019	SY2.2, SYS.3, **SYS.5**, SWE.1, SWE.2, SWE.3, **SWE.4, SWE.5, SWE.6**, MAN.3, SUP.1, SUP.8, SUP.9, SUP.10
Engine cooling	10	Europe	2020	SY2.2, SYS.3, **SYS.4, SYS.5**, SWE.1, SWE.2, SWE.3, **SWE.4, SWE.5, SWE.6**, MAN.3, SUP.1, SUP.8, SUP.9, SUP.10
Closures	7	Europe	2020	SYS.1, SY2.2, SYS.3, **SYS.4, SYS.5**, SWE.1, SWE.2, SWE.3, **SWE.4, SWE.5**, **SWE.6**, MAN.3, MAN.5, SUP.1, SUP.8, SUP.9, SUP.10, SPL.2
Power Train	15	Europe	2018	SY2.2, SYS.3, **SYS.4, SYS.5**, SWE.1, SWE.2, SWE.3, **SWE.4, SWE.5, SWE.6**, MAN.3, SUP.1, SUP.8, SUP.9, SUP.10
Driving Assistance	10	Europe	2020	SY2.2, SYS.3, SWE.1, SWE.2, SWE.3, **SWE.4, SWE.5, SWE.6**, MAN.3, SUP.1, SUP.8, SUP.9, SUP.10, ACQ.4
Engine cooling	20	Asia	2020	SY2.2, SYS.3, **SYS.4, SYS.5**, SWE.1, SWE.2, SWE.3, **SWE.4, SWE.5, SWE.**6, MAN.3, SUP.1, SUP.8, SUP.9, SUP.10
Instrument Cluster	7	Europe	2020	SWE.1, SWE.2, SWE.3, **SWE.4, SWE.5, SWE.6**, MAN.3, MAN.5, MAN.6, SUP.1, SUP.4, SUP.8, SUP.9, SUP.10, SPL.2
body electronics	15	North America	2019	SY2.2, SYS.3, **SYS.4, SYS.5**, SWE.1, SWE.2, SWE.3, **SWE.4, SWE.5, SWE.6**, MAN.3, SUP.1, SUP.8, SUP.9, SUP.10
Driving Assistance	50+	Asia	2019	SY2.2, SYS.3, **SYS.4, SYS.5**, SWE.1, SWE.2, SWE.3, **SWE.4, SWE.5, SWE.6**, MAN.3, SUP.8, SUP.9, SUP.10