VITE: Virtualisation of the Test Environment

Test process framework

Issue	2.0	Date	31/01/2019
Number of pages	21	Classification	PU

Document Reference

Project	Work package	Partner	Nature	Number
VITE	WP2	INECO	Deliverable	2.2

Partner reference (optional)

Responsible	Name/Company	Signature	Date
Author	All WP2 participants		
WP Leader	INECO		
Project coordinator	INECO		
S2RJU Project Officer	Lea Paties		

Horizon 2020 European Union Funding for Research & Innovation

Project funded by the S2R JU

S2R-OC-IP2-02-2015 – IT virtualisation of testing environment

Grant Agreement: 730815 - VITE

Shift2Rail

DOCUMENT CHANGE LOG

Issue	Date	Affected Sections	Comments
0.1	22.02.2017	All	Index and guidelines first draft for revision
0.2	18.04.2017	All	First draft for review
0.3	28.04.2017	All	Updated based on comments from partners
1.0	30.04.2017	All	Final review for delivery
2.0	31.01.2019	All	Update to include Appendix A

TABLE OF CONTENTS

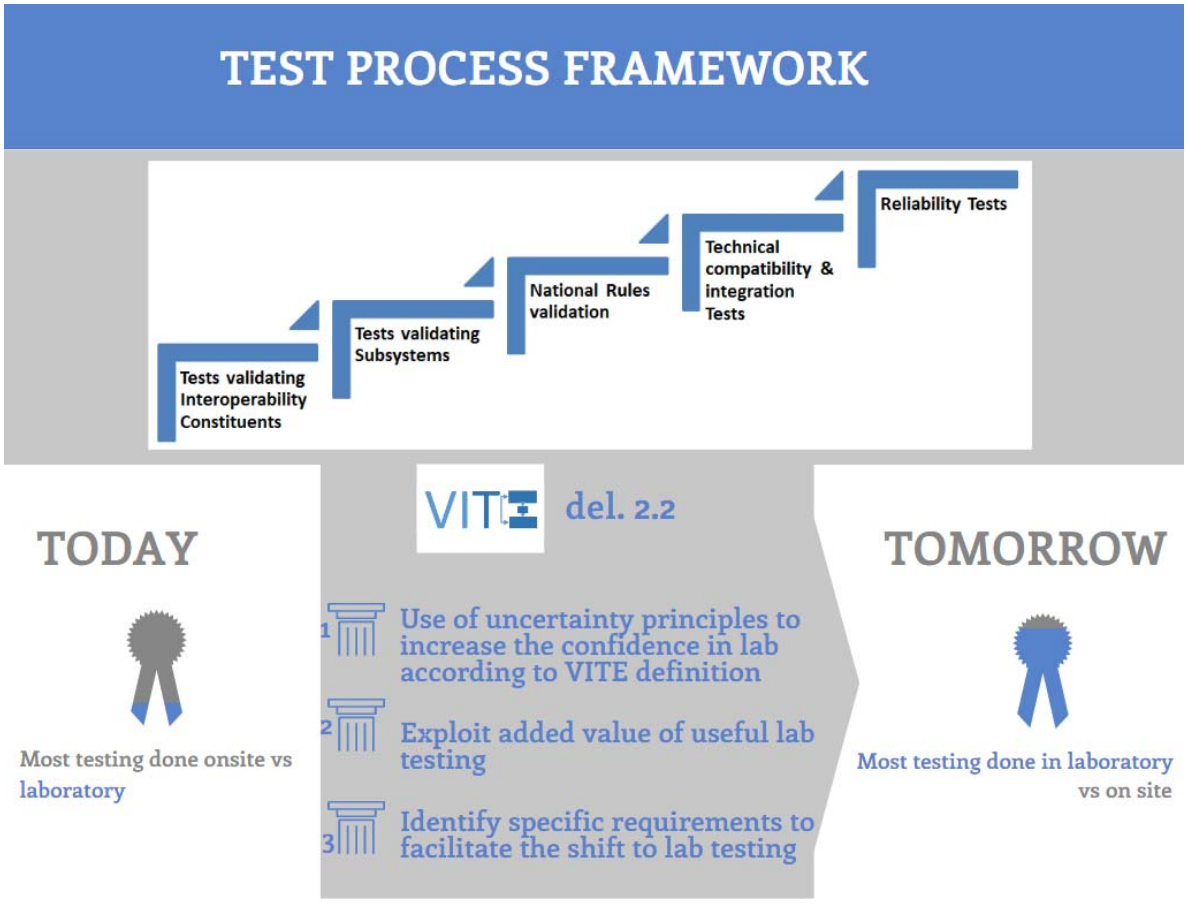
LIST OF TABLES

LIST OF FIGURES

0 EXECUTIVE SUMMARY

This deliverable describes the aspects deemed necessary by the VITE partners to contribute in a comprehensive CCS Test Process Framework to achieve the main objective of our project, i.e. to shift testing from on site to the laboratory. Therefore, it is not aimed to develop a comprehensive ERTMS test process framework but to cover the constructive approaches that could then be included in a comprehensive Test Process Framework, the CCS specifications or applicable guidelines within the Authorisation process.

In this sense, through the different inputs of the VITE partners, three main pillars have been identified to support the shifting from onsite testing to laboratory testing.



The description and contribution of all the VITE partners to these 3 pillars is the main content of this deliverable 2.2. As a consequence, this deliverable has delivered specific inputs to WP3 and WP4 of the VITE project. In addition, in order to validate these 3 main pillars.

1 INTRODUCTION

1.1 Purpose

The main objective of task 2 is to define a test process framework that will support the **shift of testing from site to lab** optimising the test case coverage model and exchange of information.

To be able to reach this objective, the needs have been assessed and identified. These include: increasing the confidence in the lab testing results on the basis of solid supporting data, simplifying the certification and authorisation process without decreasing the level of safety and allowing for a more widely accepted testing process by the rail stakeholders.

This deliverable describes the necessary issues to be included in the Test Process Framework in order to achieve the main objectives and needs identified.

1.2 Intended audience / Classification

This document is public.

1.3 Associated documentation

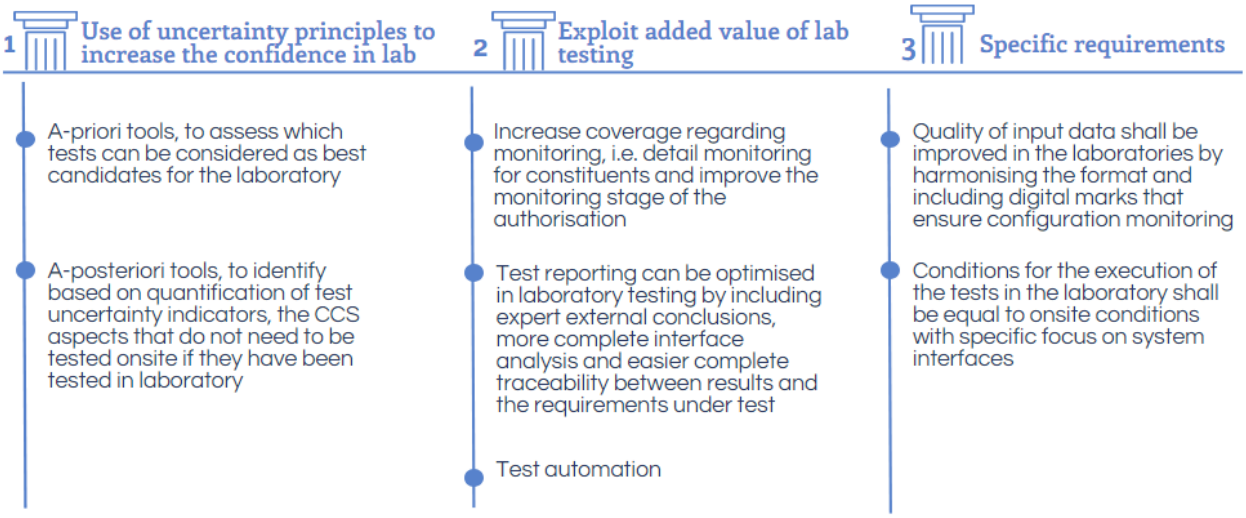
References included within each section

1.4 Abbreviations and Acronyms

CCS	Control, Command and Signalling
EC	European Commission
EVC	European Vital Computer
FS	Full Supervision mode
IPR	Indicator a Priori
IPO	Indicator a Posteriori
MA	Movement Authority
NoBo	Notified Body
NSA	National Safety Authority
S2RJU	Shift2Rail Joint Undertaking
VITE	Virtualisation of the Test Environment
OTS	Operational Test Scenarios
PC	Project Coordinator
RBC	Radio Block Centre
RFU	Recommendation For Use
SoM	Start of Mission
SB	Stand By mode
SH	Shunting mode
TSI	Technical Specification for Interoperability
WP	Work Package

2 CORE OF THE TEST PROCESS FRAMEWORK

This deliverable will contribute to the overall CCS test process framework with the following 3 pillars that based on the experience of the VITE partners will facilitate the shift from onsite to laboratory testing:



These main pillars include the aspects considered as most useful by the Infrastructure Managers, Railway Undertaking, NoBo, engineering firms, university and laboratories.

2.1 Use of uncertainty principles to increase the confidence in the laboratory

The objective of this part of the Test Process Framework is to provide a means to assess which tests can be considered as candidates for the laboratory rather than on-site, on the basis of the "confidence-building information" that is available.

The tests addressed here are seen as a part of assessment processes. The objects of these processes are CCS interoperability constituents and subsystems. There are pass-fail criteria associated with these processes. The criteria are used to judge the conformity of interoperability constituents with the TSI or whether an 'EC' declaration of verification may be granted for a subsystem.

Confidence in the results of the assessment processes is closely linked to assessment accuracy. Accuracy is a qualitative concept[1] which is quantified by uncertainty. For the types of tests addressed in VITE, with the current state of the art it is impossible to provide a rigorous definition of assessment uncertainty. The test accuracy framework described below is therefore based on the quantification of indicators of uncertainty.

Section 3 of this deliverable includes the description of the VITE Test Accuracy Framework and the practical implementation that is foreseen to be initiated during the VITE project in order to validate this pillar.

[1] ENV 13005:2000 "Guide to the expression of uncertainty in measurement. (GUM)", EURACHEM / CITAC Guide CG 4, "Quantifying Uncertainty in Analytical Measurement", Second Edition, 2000.

The uncertainty methodology that has already been used for other railway subsystems is based on 2 methods that can strongly contribute with the confidence in the selection of the laboratory to execute the tests:

- A-priori tools, with which the single causes of test uncertainty are quantified by means of uncertainty indicators and their influence on the final test result (laboratory or on-site) evaluated;

- A-posteriori tools, with which the final test result data (laboratory or on-site) are used to quantify other kinds of uncertainty indicators.

  Through the application of the above methods, and subsequent consistency checks of the results obtained with both, information will be obtained to support the identification of the most suitable testing campaigns to be executed in the laboratory based on the experience and lessons learnt of the different partners through quantitative indicators related to the specific functionalities and their results in the laboratory and on-site.
  In fact, it is clear to the different actors involved in the authorisation process that there are tests for which, without doubts, the laboratory is more suitable (e.g. subset 76), whereas there are tests that need to be executed on site (e.g. tests under operational conditions). These tools will support the decision to shift to the laboratory the rest of the tests that are in between these two limits.
  Onsite repetition of tests applicable to a specific functionality that had already been tested in the laboratory is considered to be one of the main contributing factors, in the fact that for some projects the effort and time consumption of onsite tests is 30% of the overall of the project.
  The method proposed will quantify the uncertainty of these laboratory test results based on a set of data that will be translated into measurable indicators. The first set of data will be provided by the partners of VITE and has been included in Appendix A of this report.

In general, the practical application of the Test accuracy framework is expected also to identify requirements for WP3 regarding input and output data, test criteria, time attributes (duration) and simulation of GSM-R QoS event attributes (generation mechanisms and repetition). Also for WP4 it is expected to contribute to monitoring methods for data flows to be exchanged and managed during deployment of Demo Tests/Cases Scenarios and comparative assessment for systematic cross-check between uncertainty level on-site vs. virtual lab procedures results.

2.2 Exploit the added value of the laboratory testing

It is concluded by the VITE partners that the shift to laboratory testing will be encouraged if the advantages and added value of such testing campaigns are exploited.

In this sense, several of the partners have pinpointed the monitoring aspects of the test framework as the main critical aspects that would simplify the certification and authorisation process without decreasing the level of safety:

- Firstly, it is clear that the monitoring of a bigger number of interfaces provides a major added value during a testing campaign. This monitoring has to be performed before and after each of the ETCS constituents to identify the source of any incidence.
  Specific requirements could be set to facilitate a harmonised process for the monitoring of the results that could be included in the specifications and based on the already harmonised interfaces of the system, see also section 2.3 of this deliverable
  This monitoring of different interfaces can strongly gain from the combination of the information obtained through already existing tools obtaining the information from the harmonised interfaces and protocols.

- As seen in the aviation industry, where virtual testing has various applications, including validation of technology integration potential, design, even investigation, but most importantly certification and structural reliability, number of added values are already accepted and expected. There are obvious benefits of using virtual laboratory testing, such as reduction of costs of validation and verification (often less – sometimes significantly less – than one eighth in total when comparing with an on-site test, including manufacturing), reduction of time necessary for the testing and thus of “time to market”. This applies doubly to industries that make unique one-of-a-kind designs or products (that includes complex infrastructure) and cannot easily perform many full scale test like industries producing many copies of their product, e.g. the automotive industry, and thus have to rely on analyses based on previous on-site tests. While certain specific areas, like material behaviour or some failure criteria can pose a challenge and reduce confidence in reliability of a lab test result, simulation quality improves every day through experience and new methods. Test methods can also be validated through full scale on-site tests, which would be too expensive to test “to failure” or combinations of unlikely scenarios.

- The stage for which the different partners foresee that more involvement of the laboratory will provide bigger impact is the authorisation process, mainly due to the flexibility and lack of harmonised testing campaigns. These monitoring tests are performed in some cases shortly before the authorisation and in other cases shortly after the authorisation. In most cases, the involvement of the laboratories for the execution and analysis of the data, or mainly for the analysis of the data, is considered to provide a big added value to the final commercial operation of the system.

A key aspect to achieve the shift towards lab tests is the acceptance by the stakeholders of the testing results performed in the lab. In this sense, a complete and clear testing report is critical. It is considered that test reporting from laboratory testing campaigns can be optimised for the following main reasons:

- It is easier to monitor and document the ‘observables’ from a wider range of harmonised interfaces
- It is easier due to the testing environment to trace the testing protocols and results with the requirements of the tests. These test requirements include requirements for the system under test and requirements for the execution of the tests themselves (e.g. initial conditions, etc).

It should be highlighted that no specific testing protocol is proposed within this general test process framework since it is essential that the test reporting aims to cover the objectives and means of the specific testing campaign (e.g. the traceability, requirements or format should not be the same if the system under test is an EVC constituent or the integration of the ERTMS equipped vehicle in a network).

In addition, it has to be highlighted how important the added value of automated laboratory test execution is. It is clear that minimising the effect of human error for the testers will contribute in the acceptance of the test results by all stakeholders. This input will be taken into account by WP3.

Finally, on-site testing is often too long because several iterations of the tests with different SW versions have to be performed. Test in laboratory are optimum to debug the system while maintaining a correct traceability of the SW configuration management of these changes. The main advantages of the lab testing in this sense is that there can be more intensive tests to debug the errors of the ERTMS system and their integration with the collateral systems that are included in the laboratory (e.g. the rail operating centers, interlockings, etc) they can be performed in shorter periods

of time and the updates of the system under tests can be made easier to correct the detected bugs short after their identification. On site, and mainly for lines and vehicles in commercial service, this flexibility is never possible. The time and economic effort necessary to debug the system can therefore be reduced considerably when all the test phase for debugging the system is performed in the laboratory and then the unavoidable few test scenarios performed on site can be executed with the perfectly stable version of the ERTMS system that will be placed in service. In addition, there are specific activities that have to be performed before any testing onsite that do not need to be performed for the laboratory testing, mainly safety analysis of the integration of intermediate versions of the system under tests can be delayed until the end of the debugging tests in the laboratory whereas for onsite testing, the safety aspect of the system has to be finalised even of the testing would be of intermediate versions of the system under test. Lab testing should be used in the first phases of testing to reduce to a minimum on-site testing to the demonstration of the final SW application.

2.3 Identify minimum requirements to be included in specifications and guidelines

During the work of this deliverable, specific requirements have been identified that if included in legislation guidelines or the comprehensive test process framework will facilitate the shift to laboratory testing.

- It is critical to improve the quality of the input data that is delivered to the laboratory and the process followed. A comprehensive and digital tool for configuration management would increase the acceptance of the laboratory test results.

  WP 3 will analyse if such an input for a specific tool or input data format (e.g. analogue to a digital signature for the validation of the data) can also be included in the scope of the lab architecture.

  A proposed requirement is the following:

  'Input data and configuration parameters for the execution of the tests shall be according to the latest validated data by the supplier. In addition, this shall be supported by the digital signature of this data by the supplier and include the specific version identification within the Configuration Management Plan or release notes'

- It is essential that conditions for the execution of the tests in the laboratory be equal to onsite conditions. WP3 will also take into account these proposals as input to their work.
  '*The equipment tested in the laboratory shall be the real equipment installed onsite. The additional equipment necessary to perform the tests shall be as similar as possible to the systems as installed onsite, this can include virtual equipment if demonstrated that the conditions are analogue to the onsite conditions'*

  The following specific requirements have been identified as most relevant:
  - For the overall test framework, laboratories should provide a document with the technical limitations. To increase the confidence in the laboratory testing, these limitations should be zero or very small:

    '*Before the execution of any testing campaign, the specific laboratory shall present a document containing the technical limitations associated to the execution of the tests. This document shall be annexed to the Test report'*

  - The observables should be measured in harmonised interfaces or open protocols to ensure the confidence of the stakeholders in the results and should include at least the registers available in on-site testing:
    '*Test results shall be comparable and therefore need to be identified in the harmonised interfaces'*

- The laboratory test report should include not only what has been tested but also how it has been tested to increase the confidence in the results. These include the configuration information as mentioned before (e.g. SW release notes), numbers of tries in a test, traceability to the requirements, etc.

- All the actions that build these three pillars can contribute if taken into account within the existing legislation, the comprehensive CCS testing Framework and existing guidelines related to CCS authorisation. For example, it is recognised by the partners of VITE that it would be valuable to build based on the tools, principles and requirements included within the pillars, a clear guideline for NoBos on the classification of all test campaigns (if they are required to be tested on-site or could also be accepted from the laboratory) for which the reports are evaluated within the certification works (i.e. those as included in the CCS TSI). The document will be developed initially by the 3 NoBos involved in VITE but also questionnaire or opinion for rest of NoBos can be sought outside the project. During WP5 it will be studied if any of the conclusions reached within the VITE project will derive in an RFU in NBrail to cover the results obtained of this analysis.

3 UNCERTAINTY METHODOLOGY APPLIED TO CCS

3.1 VITE Test Accuracy Framework

The Test Accuracy Framework is a part of the Test Process Framework providing a means to assess which tests can be considered as candidates for the laboratory rather than on-site, based on the quantification of test uncertainty indicators.

It is based on the TrioTRAIN Uncertainty Framework[2], which was applied to three different assessment processes - two associated with 'EC' verification of subsystems (rolling stock aerodynamics and running dynamics) and one associated with conformity of an interoperability constituent (pantograph).

In terms of accuracy considerations, the main difference between the TrioTRAIN and the VITE specific cases are the nature of the parameters used in the assessment process. The parameters are continuous random variables for TrioTRAIN (e.g. air pressure measured alongside the track, derailment ratio, pantograph-catenary contact force) and Boolean random variables for VITE (see below for more details). This has required slight reformulations of the concepts identified in TrioTRAIN. The structure of the framework, comprising five elements (definitions, objects, parameters, methods, references) is thus re-described below as tailored to the needs of the VITE project.

3.1.1 Definitions

The definitions below are useful for the VITE project. They are taken from a variety of sources and adapted, or created, for the purposes of the project.

The VITE Test Accuracy Framework embraces the concepts that are behind international standards such as ISO 5725-1:1994[3], ISO Guide 98-3: Guide to the expression of uncertainty in measurement (GUM:1995)[4]. In these documents, it is recognised that in many practical cases repeated experiments are not possible, and methods are given on how to quantify accuracy even in those cases.

accuracy (schematic definition, adapted from ISO description[5]) = precision + trueness

uncertainty (schematic definition): quantification of accuracy

uncertainty of a measurement (GUM): "*parameter, associated with the result of a measurement, that characterises the dispersion of the values that could reasonably be attributed to the measurand*"

precision (ISO description): closeness of agreement between test results (same test, same test object)

example: an assessment is characterised by high precision if for a same system under test the same result is obtained in a high number of repeated assessments

trueness (ISO description): closeness of agreement between the arithmetic mean of a large number of test results and the true or accepted reference value

---

[2] Licciardello, R., Funfschilling, C., & Malavasi, G. (2016). Accuracy of the experimental assessment of running dynamics characteristics quantified through an uncertainty framework. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 0954409716657373.

[3] ISO 5725-1:1994, Accuracy (trueness and precision) of measurement methods and results -- Part 1: General principles and definitions.

[4] ISO/IEC Guide 98-3:2008 Uncertainty of measurement -- Part 3: Guide to the expression of uncertainty in measurement (GUM:1995)

[5] www.iso.org/obp/ui/#iso:std:iso:5725:-1:ed-1:v1:en

example: an assessment is characterised by high trueness if a system that is actually conform is judged to be conform in a high number of repeated assessments because the mean of the assessment results is very close to unity (=pass, true)

repeatability conditions (ISO 3534-1): conditions where independent test results are obtained with the same method on identical test items in the same laboratory by the same operator using the same equipment within short intervals of time

reproducibility conditions (ISO 3534-1): conditions where test results are obtained with the same method on identical test items in different laboratories with different operators using different equipment

repeatability (ISO 3534-1): precision under repeatability conditions

reproducibility (ISO 3534-1): precision under reproducibility conditions

indicator of uncertainty (TrioTRAIN --> VITE): quantitative parameter providing information on accuracy when uncertainty itself is not practically quantifiable according to its definition; precision indicators and trueness indicators are two types of indicators

example: ratio of the number of 'fail' / 'pass' under repeatability conditions (precision indicator); result of a test on a reference test object i.e. that is known to be e.g. conform (trueness indicator)

assessment process (TrioTRAIN --> VITE): test process following specific protocols/procedures, with the objective of judging whether the test object meets specific pass/fail criteria

examples: operational test scenarios (TSI §6.1.2), assessment procedures for CCS interoperability constituents (TSI §6.2.1), assessment procedures for CCS subsystems (TSI §6.3)

assessment quantity (TrioTRAIN --> VITE): single (Boolean) variable that must meet a specific pass/fail criterion as a necessary condition for the overall assessment to be passed

examples: ‘RBC1 sends announcement to RBC2’, ‘establishment of communication session EVC-RBC2’, ‘EVC sends position report to RBC1, RBC2’…

assessment value (TrioTRAIN --> VITE): value of the assessment quantity (e.g. for a Boolean quantity, true or false)

assessment uncertainty (TrioTRAIN) = the uncertainty associated with the assessment quantity that is being assessed, represented, once the assessment method is proven, by the assessment-to-assessment variability.

examples: ‘RBC1 sends announcement to RBC2’ – assessment uncertainty occurs e.g. if two different protocols, or two different assessors applying the same protocol, would attribute a different pass/fail result to this variable; this example is more relevant for test conditions that are ambiguous (which is not the case for the example assessment quantity given above) and could lead to different results on the part of different assessors

acceptable assessment uncertainty (TrioTRAIN --> VITE): assessment uncertainty that is considered safe and effective; for railway applications we can consider what has been done in the past to be safe (Safety Directive art. 4), thus as a starting point the uncertainty associated with tests that have been accepted can be considered as acceptable; for the case of VITE, we will consider on-site tests as the reference against which to measure laboratory tests

false pass (VITE): the result of a test for a given assessment quantity that should have given fail because the test object is non-conform, but gave pass due to lack of accuracy of the test itself

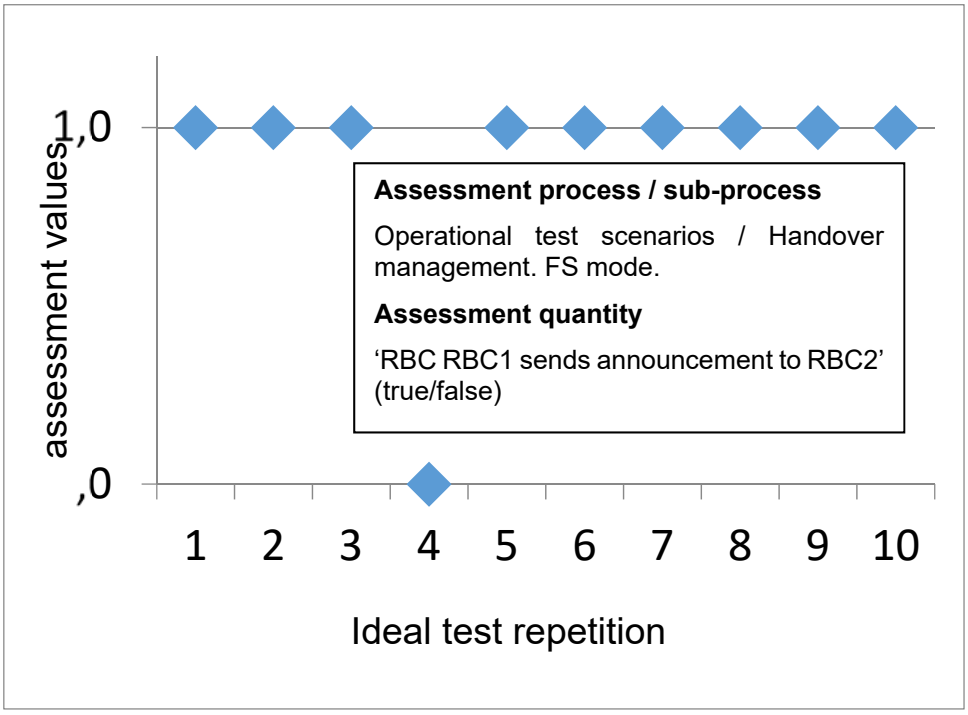


Figure 1. Illustration of precision of an assessment process for a particular assessment quantity. Test n. 4 has given a (presumably false) 'fail' e.g. due to ambiguous definition of the assessment quantity.

3.1.2 Objects

The objects addressed by the VITE Test Accuracy Framework are the technical assessment processes identified in TSI CCS - operational test scenarios (TSI §6.1.2), assessment procedures for CCS interoperability constituents (TSI §6.2.1), assessment procedures for CCS subsystems (TSI §6.3).

The objects of the framework are not to be confused with the test objects, which are in the case of VITE the devices subjected to the tests defined in the above assessment processes.

3.1.3 Parameters

The following figure illustrates the hierarchy considered in the Framework, using as an example the assessment process "operational test scenarios" (TSI §6.1.2). The hierarchy consists of a top level (assessment processes), an intermediate level (scenarios, or sub-processes) and a bottom level (scenario elements or equivalently assessment quantities). All of the elements of the hierarchy may be considered as "parameters" of the framework in the sense used in TrioTRAIN. They are Boolean variables, assuming the two values 'true' / 'false', or 'pass' / 'fail'. For the intermediate level to be a 'pass', all corresponding bottom level elements have to be 'pass'. For the top level to be a 'pass', all intermediate level elements have to be 'pass'.

The lower level parameters are also the elementary assessment quantities (i.e. the ones that have to meet a pass/fail criterion), that assume two assessment values ('true'/'false').

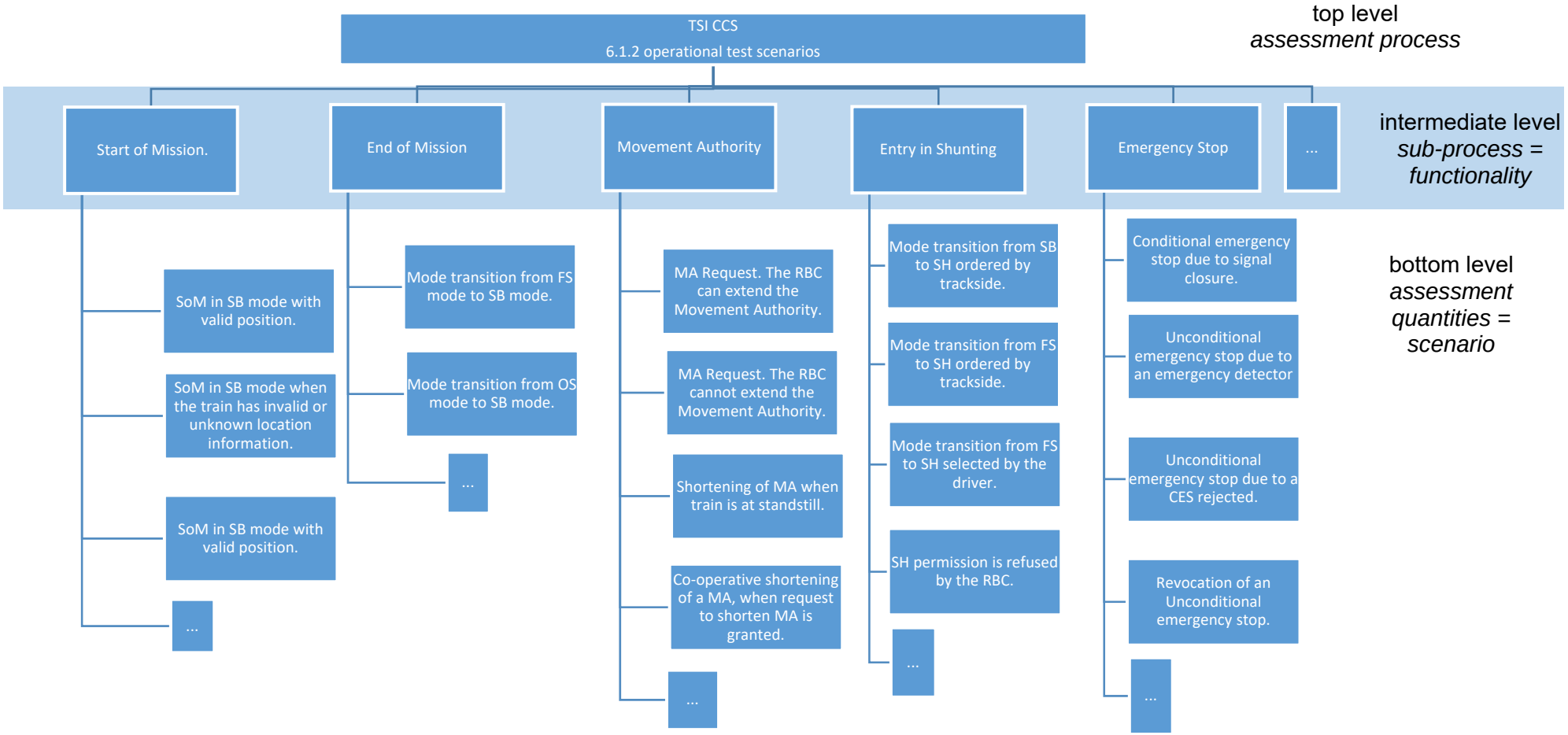


Figure 2. Structure – assessment process, sub-process, assessment quantities.

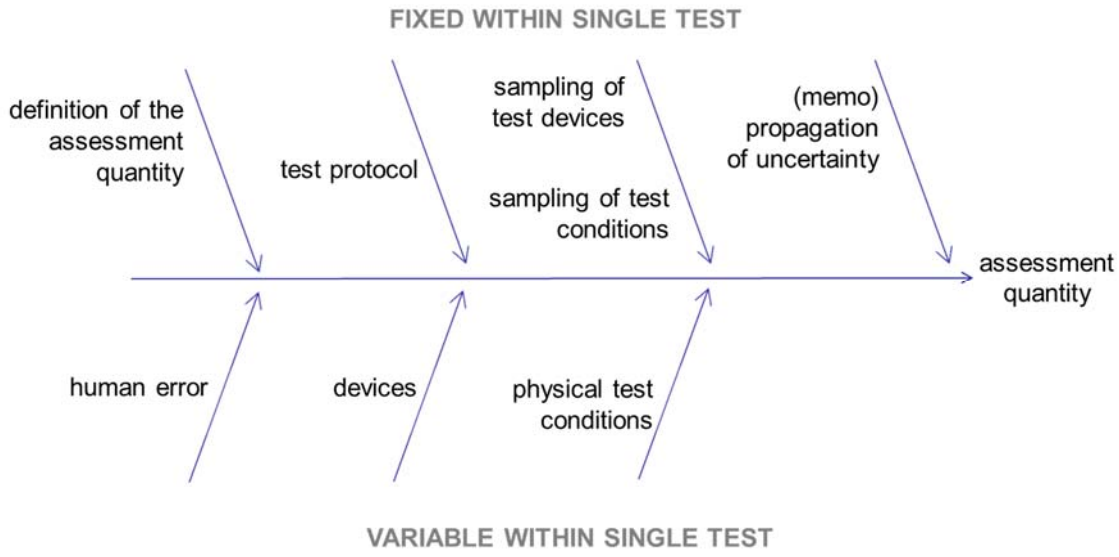
3.1.4 Methods

There are essentially two possible approaches to quantify uncertainty[6]:

- the a-priori approach, in which the causes of uncertainty are identified, linked with the assessment quantity, and individually quantified so as to be able to quantify their contribution to the overall assessment uncertainty
  example of a-priori analysis: the two main uncertainty causes identified for a test are the possibility of the instrument in use giving a false reading (1/1000 cases) and the assessor not recognising a false reading (1/10 cases); the two causes are independent and both need to occur simultaneously in order to generate a false result, thus the overall accuracy may be quantified through the product of the two probabilities, i.e. the test could lead to 1/10,000 cases giving a false result
- the a-posteriori approach, in which the causes of uncertainty are not considered, and the quantification of assessment uncertainty is gained through the knowledge of test results
  examples of a-posteriori analyses: 5 repeated tests for the same assessment quantity **on the same object** under repeatability conditions have given 5 times the same result; 5 repeated tests for the same assessment quantity under repeatability conditions have given 5 times the same result; two different indicators are used in these examples, giving different degrees of confidence regarding the accuracy of the test; often for railway applications this kind of information is the only one available, and often the numbers are quite few, still any information is better than none.

The above approaches may be implemented in different ways, from very simple techniques to very sophisticated statistical and probabilistic analyses. In this report we start from the simplest techniques.

With the a-priori approach, the first step is the identification of the sources of uncertainty and how they are linked. A tentative graphical representation for the tests within the scope of VITE is given in Figure 3, and the corresponding descriptions are given in Table 1.



[6] Licciardello, R., Funfschilling, C., & Malavasi, G. (2016). Accuracy of the experimental assessment of running dynamics characteristics quantified through an uncertainty framework. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 0954409716657373. Licciardello, R., Grappein, E., & Rueter, A. (2015). On the accuracy of the assessment of open-air pressure loads due to passing trains: Part 1: Experimental assessment. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 229(6), 644-656.

Figure 3. Fishbone diagram representing the main sources of uncertainty for the assessment quantities addressed in VITE.

Table 1. Description of the main uncertainty sources of Figure 3 and comments on their entity in laboratory and in on-site tests.

Uncertainty source	Laboratory vs on-site testing
uncertainty due the definition of the assessment quantity • some assessment quantities could have ambiguous definitions, leading to different possible outcomes depending on the assessor	similar for both lab / on-site
uncertainty due to the test protocol • the test protocols leave some requirements open, to the choice of the applicant; if tests are ideally repeated with these elements left to vary between the possible choices, a number of different outcomes could be possible • when different devices (instruments, interfaces) are available to provide the same information, the test protocol could require proof from a device that is not suitable for this purpose	possibility to perform lab tests with more options than on-site tests
uncertainty due to the devices (instruments, interfaces) used to measure the assessment quantity • the device measuring the assessment quantity e.g. the 'RBC1 sends announcement to RBC2', could give a false reading	similar for both lab / on-site
uncertainty due to human error in using the devices or in recording results • the required information could be derived from the wrong device • a pass result could be recorded when the actual results was a fail	similar for both lab / on-site; potential for higher probability of error under the more demanding work environment of on-site tests
uncertainty due to sampling • tests are performed generally on small "samples", i.e. in specific test conditions and only on representative objects (e.g. test vehicle, test wayside devices); a component of uncertainty is thus linked to the possible events that could occur under the different conditions and with the different objects that are actually never tested	possibility for lab tests to consider potentially broader (and controlled) test conditions and more test objects
uncertainty due to variability of physical conditions during tests • some physical quantities are variable during tests and there is also an associated assessment-to-assessment variability for these - e.g. train speed, braking intensity, temperatures (ambient, devices, etc.), vibrations; it is not guaranteed that the test would give the same result under different conditions that are not tested (e.g. with a very high temperature of a certain device)	on-site tests are generally more demanding in this sense, although they cannot ensure that all unfavourable conditions occur; lab tests have the

	potential to impose severe conditions
propagation of uncertainty • this is a source of uncertainty that occurs with mechanical measurements for example (TrioTRAIN applications); it is kept here as a memo in case it is identified as a source also for the VITE scope	-

With the a-posteriori approach, test results are necessary for the analysis. Possible analysis techniques are:

- comparison of laboratory results and on-site results for the same assessment quantity;
- counting number of 'fail' and 'pass' for different assessment quantities under different conditions;
- Bayesian techniques to incorporate the evidence obtained through available test results in confirming hypotheses on the degree of accuracy of the test.

3.1.5 References

The references to documents addressing the definitions, objects, parameters and methods characterising the VITE Test Accuracy Framework are given, as an initial list to be further populated as the framework evolves, in the above sections.

3.2 Practical application of the Test Accuracy Framework

Based on the definitions of the framework, a test on the CCS subsystem is considered to be accurate when, if performed on an object that is conform to the TSI, it has an extremely low probability of giving 'fail' as a result and, vice-versa, if performed on an object that is not conform to the TSI, it has an extremely low probability of giving 'pass' as a result.

For the purposes of VITE, it is interesting to examine the accuracy of laboratory tests versus the accuracy of the corresponding on-site tests. This activity will be initiated within the VITE project, based on this framework, with an uncertainty analysis aiming at quantifying suitable uncertainty indicators and assessing their acceptability.

Example. A possible statement of uncertainty (or accuracy) for a test is, "the accuracy of the test is such as to have no more than 1/1000 false 'fail' and no more than 1/1000 false 'pass'". The indicators used to quantify uncertainty are in this case the number of false 'fail' and the number of false 'pass'.

The analysis may be performed with both a-priori and a-posteriori approaches. It will be seen during the continuation of VITE which of the methods described above, or other ones identified, are the most appropriate, based on the data available. At the end of the activities, recommendations will be made as to how to continue to populate the framework as continuous support to the cause of transferring tests to the lab where safe and effective.

3.2.1 Indicators for the a-priori approach

The a-priori approach considers the cause of uncertainty in tests results and ideally quantifies their effects on the final test result, depending on the available data.

Within the VITE project, based on the data available to the partners, it would be interesting to begin quantifying the indicators in the following non-exhaustive list (linked to the uncertainty sources of Figure 3):

- IPR1 n. of known occurrences of ambiguous definitions in test protocols leading to false recordings

- IPR2 n. of known occurrences of human error (description and possibly differences between on-site and lab testing)
- IPR3 n. of known different results due to the application of different (but equivalent) protocols
- IPR4 n. of known occurrences of errors due to devices (human error, or device giving the wrong information)
- IPR5 n. of known occurrences of physical test conditions affecting the result of the test (e.g. different results depending on ambient temperature etc.)
- IPR6 n. of known occurrences of differing test results due to tests performed on different devices of the same type

Available test protocols in VITE should be analysed in a view to identify the above sources (even in the absence of actual data) and the possible differences related to on-site and lab testing.

3.2.2 Indicators for the a-posteriori approach

The a-posteriori approach quantifies uncertainty by comparing test results without going into the detail of uncertainty sources.

Within the VITE project, based on the data available to the partners, it would be interesting to begin quantifying the indicators in the following non-exhaustive list:

- IPO1 for each sub-process (or even each assessment quantity if feasible), for available lab tests and on-site tests: ratio of ‘fail’ to total number of tests
- IPO2 for each sub-process (or even each assessment quantity if feasible), for available lab tests and on-site tests: ratio of false ‘fail’ to total number of tests
- IPO3 for each sub-process (or even each assessment quantity if feasible), for available lab tests and on-site tests: ratio of false ‘pass’ to total number of tests

When these indicators are obtained for several test campaigns, the sub-processes where the indicators are favourable can be identified (e.g. low ratio of false ‘fail’ to total number of tests). These cases are candidates for shifting testing to the laboratory only.

Another set of uncertainty indicators are related to cases in which both laboratory and on-site tests were performed for the same sub-process (or assessment quantity). In order to identify those for which lab testing would deserve high confidence due to past experience, it would be interesting to quantify the number of occurrences of the cases listed in Table 2.

Table 2. Cases of interest related to existing lab / on-site test information.

Lab test	On-site test	Actual operations	Description/comments	N. of occurrences (tbc during VITE)
×	✓	✓	On-site test did not identify problems, lab test did (e.g. due to a higher number of repetitions)	...
✓	✓	×	Problems identified only during operation, information to be drawn e.g. from safety alerts on EUAR web-site	...
✓	×	✓	On-site test identified problems that were corrected and actual operations worked out well	...
✓	×	×	Lab test did not identify any problem, but on-site tests and operations were not satisfactory	...

…	…	…	other cases of interest	...

3.2.3 Use of the uncertainty indicators as confidence-building information

The collection of data regarding the uncertainty indicators, for example according to Table 3, will begin to populate the framework with information that has the potential of building confidence on e.g.:

- which assessment quantities are suitable for lab tests only;
- which assessment quantities absolutely require on-site tests;
- for which assessment quantities can the availability of lab test results allow the simplification of on-site tests
- etc.

Table 3. Template for collecting uncertainty indicator information.

		UNCERTAINTY INDICATOR																	
		IPR1		IPR2		IPR3		IPR4		IPR5		IPR6		IPO1		IPO2		IPO3	
Sub-process	Assessment quantity	Lab	On-site	Lab	On-site	Lab	On-site	Lab	On-site	Lab	On-site	Lab	On-site	Lab	On-site	Lab	On-site	Lab	On-site
Handover management. FS mode.	RBC1 sends announcement to RBC2	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Handover management. FS mode.	establishment of communication session EVC-RBC2	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Handover management. FS mode.	EVC sends position report to RBC1, RBC2	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

The detailed methodology for this purpose has to be developed on a case-by-case basis. Starting from the simple approaches, just the identification e.g. of assessment quantities for which the partners have performed several lab tests and also corresponding on-site tests, and the results were always consistent, would support the use - for the assessment quantity in question - of lab testing only. Such cases, if existing, would be "low-hanging fruit" for the VITE project.

In the more likely event of no simple cases, the study will have to go more in depth with more exploratory data analysis and the use of statistical techniques. An example could be the use of a Bayesian approach for the quantification of uncertainty indicators based on the available evidence: for example, testing hypotheses regarding the 'fail'/'pass' ratio based on the test results used as evidence.

4 CONCLUSIONS

The main description of key improvements that VITE project considers necessary to include within a comprehensive ERTMS Test Process Framework in order to achieve a shift from on site to laboratory testing can be found in the executive summary of this deliverable.

This section includes the detailed conclusions that are proposed:

- Test accuracy framework will be further populated with data from ERTMS test reports within WP2 in order to support the identification of the most suitable testing campaigns to be executed in the laboratory. Results will be an input for WP3 mainly regarding input and output data, test criteria, time attributes (duration) and simulation and to WP4 to contribute to monitoring methods.
- To exploit the added value of the laboratory testing VITE project will focus in maximising the following advantages of laboratory testing:
    - Incorporate within lab architecture, execution and monitoring of test results the monitoring of a bigger number of interfaces provides a major added value during a testing campaign. (all VITE WP)
    - Assess, with the known aviation principles, the best options for virtualisation of the laboratory while maintaining the analogue behaviour for onsite (WP3)
    - Focus specially to improve the testing campaigns of the last stages before authorisation since they are the ones where the partners foresee a bigger possibility of shift from on site to laboratory testing (all VITE WP)
    - Optimise the laboratory test reports by incorporating all the information (mainly more observables and traceability) that is possible due to the tests being performed in the laboratory (WP2, WP4 and WP5)
    - Include possibility for automation in the laboratory architecture (WP3)
    - Maximise the advantages of lab testing for system debugging (WP3)
- The following are the requirements that have been identified within this VITE deliverable 2.2:
    - *'Input data and configuration parameters for the execution of the tests shall be according to the latest validated data by the supplier. In addition, this shall be supported by the digital signature of this data by the supplier and include the specific version identification within the Configuration Management Plan or release notes' (WP3)*
    - '*The equipment tested in the laboratory shall be the real equipment installed onsite. The additional equipment necessary to perform the tests shall be as similar as possible to the systems as installed onsite, this can include virtual equipment if demonstrated that the conditions are analogue to the onsite conditions' (WP3)*
    - '*Before the execution of any testing campaign, the specific laboratory shall present a document containing the technical limitations associated to the execution of the tests. This document shall be annexed to the Test report' (WP3)*
    - '*Test results shall be comparable and therefore need to be identified in the harmonised interfaces' (WP3)*
- Improve configuration management possibilities and traceability for the laboratory testing (all WP)
- WP5 evaluate the possibility to perform a clear guideline for NoBos on the classification of all test campaigns, if they are required to be tested on-site or could also be accepted from the laboratory

END OF DOCUMENT

VITE: Virtualisation of the Test Environment

Appendix A to the Test process Framework: Test Accuracy Methodology

Issue	1.0	Date	30/10/2018
Number of pages		Classification	PU

Document Reference

Project	Work package	Partner	Nature	Number
VITE	WP2	INECO	Supporting doc	01

Partner reference (optional)
Dicea and Ineco

Responsible	Name/Company	Signature	Date
Author	Riccardo Licciardello, Cristina Pardo, Silvia Domínguez		
WP Leader	INECO		
Project coordinator	INECO		
S2RJU Project Officer	Lea Paties		

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

0 EXECUTIVE SUMMARY

ERTMS/ETCS testing according to the European regulatory framework revolves around the Control-Command and Signalling Technical Specification for Interoperability. Experience with this TSI has shown that the process to put subsystems and components into service is long and complex, and not always are problems during operation avoided. The Shift2Rail project called VITE - VIrtualisation of the Test Environment for signalling - addresses a part of this problem by contributing to the analysis of which laboratory tests can usefully replace on-site tests - the latter being the more costly and complex - without affecting the high level of safety achieved.

Experience with this kind of tests is such that the experts have opinions on which parts could usefully be performed only in the laboratory. However, at the moment there is no scientific approach to back - or confute - these opinions. The aim of this work is methodological: to introduce a test accuracy framework for this purpose, partially populated with data from VITE, showing a possible scientific approach towards the gradual replacement or integration of on-site tests with laboratory tests wherever useful. The approach is based on the use of uncertainty indicators to arrive at quantitative information on test accuracy.

The analysis has followed two approaches, based on the TrioTRAIN framework already used for other railway subsystems (rolling stock - aerodynamics and running dynamics, and pantograph-OCL interaction), with the difference that in VITE the assessment values are Boolean variables (Pass/Fail) whereas for the other cases they were continuous variables. The approaches are:

- A priori, in which the causes that brought to the test result (i.e. Pass or Fail) are studied - these are e.g. ambiguous test protocols, human error etc.
- A posteriori, in which the actual numbers of Pass/Fail with respect to the total number of tests are analysed, to understand the differences between different types of test categories (e.g. Start of Mission, RBC-RBC handover etc. etc.).

Research performed conclude that it will be difficult to reach ERTMS indicators that will allow to evaluate ERTMS related test accuracy in its academic sense. However, the results of the methodology applied during the project provide interesting results for increasing the confidence of laboratory testing.

The results of the analysis are:

- A set of ERTMS/ETCS functionality test-categories for which the analysis is proposed to continue;
- A list of identified causes of Fail results (a priori analysis)
- A list of observables for laboratory and on-site tests through which such tests may be monitored in order to collect information on the causes of lab or on site inaccuracies that can justify or not the shift of certain test categories to the laboratory only (a priori analysis)
- A list of indicators that are considered to be useful for monitoring future test campaign results
- Tables and charts which can show the relative importance of the different functionality test categories in terms of occurrence of failures and mismatches between lab and on-site test results
- A criterion for the minimum number of tests needed to achieve robust conclusions
- A possible roadmap for implementation of the framework and possibly reach the identification of test accuracy indicators
- Recommendations to improve ERTMS/ETCS test reports in support of a safe on-site →lab shift functionality by functionality.

1 INTRODUCTION

1.1 Purpose

The aim of this document is to present a way of reporting that can set the basis for increasing confidence in lab testing as a replacement for on-site testing where appropriate.

1.2 Intended audience / Classification

This document is public.

1.3 Associated documentation

References included within each section

1.4 Abbreviations and Acronyms

CCS	Control, Command and Signalling
EVC	European Vital Computer
FS	Full Supervision mode
IPR	Indicator a Priori
IPO	Indicator a Posteriori
MA	Movement Authority
NoBo	Notified Body
NSA	National Safety Authority
S2RJU	Shift2Rail Joint Undertaking
VITE	Virtualisation of the Test Environment
OTS	Operational Test Scenarios
PC	Project Coordinator
RBC	Radio Block Centre
RFU	Recommendation for Use
SoM	Start of Mission
SB	Stand By mode
SH	Shunting mode
TSI	Technical Specification for Interoperability
WP	Work Package

2 TEST ACCURACY METHODOLOGY

2.1 Test Accuracy Framework

In engineering, an important way to build up confidence in a test process is to gain information about its accuracy. In fact, accuracy means precision (repeatability/reproducibility) - i.e. accurate test processes will give similar or identical results when repeated under different conditions - and trueness - i.e. accurate test processes will provide similar results if compared with reference test results obtained through other trusted means. However, this analysis requires a large effort for complex systems and is seldom done for routine jobs.

In the VITE project, the recourse to the analysis of uncertainty - which is the quantitative aspect of accuracy - is performed with the objective of providing analytic and quantitative information regarding issues that with the current state of the art are often only provided in qualitative form (e.g. laboratory tests are more controllable, more repeatable, than on-site tests). It is useful for a research project to perform this effort and explore these aspects quantitatively, so that subsequent routine projects may benefit from the methods and results developed.

The methodology adopted in the project is represented by the VITE Test Accuracy Framework.

Research performed conclude that it will be difficult to reach ERTMS indicators that will allow to evaluate ERTMS related test accuracy in its academic sense. However, the results of the methodology applied during the project provide interesting results for increasing the confidence of laboratory testing.

The Test Accuracy Framework is a part of the Test Process Framework described in deliverable D2.2 **¡Error! No se encuentra el origen de la referencia.**. It is based on the quantification of test accuracy indicators and is intended to support replicable processes for the quantification of such indicators, applicable to both lab tests and on-site tests, so as to identify areas (in practice, ERTMS/ETCS functionalities) where accuracy is in favour of one or the other.

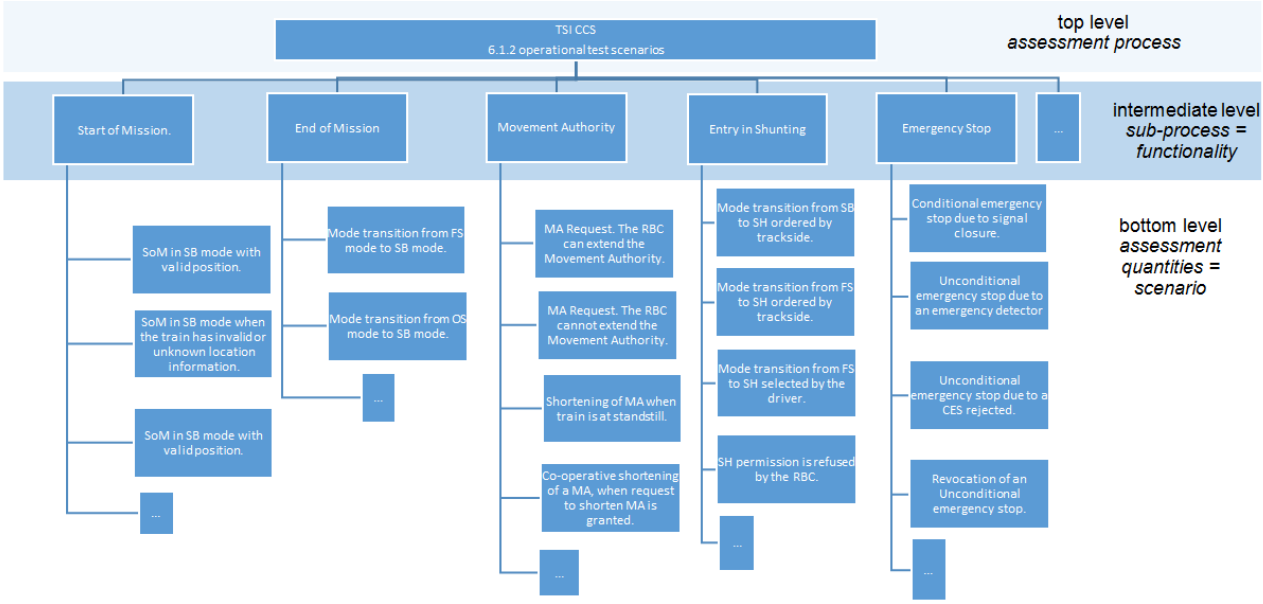
The starting point for its definition has been the Uncertainty Framework of the TrioTRAIN projects, one of them was applied to three different assessment processes - two associated with ‘EC’ verification of subsystems (rolling stock aerodynamics and running dynamics) and one associated with conformity of an interoperability constituent (pantograph). The structure of the framework, comprising five elements (definitions, objects, parameters, methods, references) has remained the same and has been tailored to the needs of the VITE project.

The objects of analysis, according to the VITE's scope, are the technical assessment processes identified in TSI CCS (Commission Regulation EU 2016/919 of 27 May 2016) - operational test scenarios (TSI §6.1.2), assessment procedures for CCS interoperability constituents (TSI §6.2.1), assessment procedures for CCS subsystems (TSI §6.3). The objects of the framework are not to be confused with the test objects, which are in the case of VITE the devices/software subjected to the tests defined in the above assessment processes.

The definitions from TrioTRAIN have required slight modifications. For VITE, the assessment quantity is a "single (Boolean) variable that must meet a specific pass/fail criterion as a necessary condition for the overall assessment to be passed". For the topic under analysis, the appropriate level identified for this purpose is the "scenario" level (see). As an example, a scenario might examine whether the start of mission procedure takes place correctly in the ETCS Stand By mode. For this to be verified, single boolean variables (the parameters) such as ‘bi-directional exchange of messages between RBC and EVC is recorded in the JRU’, ‘EVC switched from SB mode to SR mode’, ‘once the position is known, the EVC switches from SR to FS mode’ all have to be 'true' or 'pass'. If this occurs, then the scenario/assessment quantity "Start of mission in SB mode” is a 'pass'. The assessment uncertainty - i.e. as defined in TrioTRAIN "the uncertainty associated with the assessment quantity that is being assessed, represented, once the assessment method is proven,

by the assessment-to-assessment variability", is not quantified by a range of "values that could reasonably be attributed to a measurand" as for mechanical measurements, rather by an uncertainty indicator such as the probability that a non-conform system would pass the assessment (probability of a "false pass").

Figure 1 Assessment quantities as defined for the scope of the VITE project (ERTMS/ETCS) tests).



Ideally, the purpose of the analysis is to quantify the uncertainty indicators (e.g. probabilities of false passes) for the different scenarios, based on the data available to the project partners as part of their background or generated during the demonstration activities. For this purpose, there are two basic categories of **methods** in the framework:

- the a-priori approach, in which the causes of uncertainty are identified, linked with the assessment quantity, and individually quantified so as to be able to quantify their contribution to the overall assessment uncertainty;
- the a-posteriori approach, in which the causes of uncertainty are not considered, and the quantification of assessment uncertainty is gained through the knowledge of test results - e.g. the numbers regarding passes and fails for each test case.

The above approaches may be implemented in different ways, from very simple techniques to very sophisticated statistical and probabilistic analyses. In the VITE project, we start from the simplest techniques.

With the a-priori approach, the first step is the identification of the sources of uncertainty and how they are linked.

The a-posteriori approach for the start of mission scenario would be implemented, ideally, by repeating the scenario a number of times in the lab and on-site and comparing the results in terms of frequency of passes or fails. However, this is usually impracticable due to the large number of tests required, so other types of information have been sought from the partners' backgrounds (e.g. how many times has a given scenario failed? in the lab? on site? has it ever passed in the lab and failed on site?) and from the demonstration tests (how many fails of single parameters are recorded? for which scenarios? etc.). These are all indicators regarding the accuracy of the tests that can help,

on the one hand, gain confidence in the tests which prove to be solid and, on the other, identify tests that might require improvement.

2.2 Methodology

2.2.1 Steps and basic categorisation

The generic framework described above was translated into concrete steps that are described in this sub-section. The practical application is further detailed in §3 and §4 for the a-priori and a-posteriori analyses respectively.

The a-priori analysis addressed, in this work, lab and on site tests. It involved the following steps:

- Consultation of test reports;
- Consultation of test experts;
- Drafting of list of a priori indicators and of consequent observables (see subsequent section §2.2.2) - i.e. influence factors of the test result (pass/fail) that could vary during repeated tests;
- Observations of tests for the different functionality test-categories of **¡Error! No se encuentra el origen de la referencia.**;
- Analysis of the observations of the VITE demonstration tests.

The a-posteriori analysis also involved both lab and on-site tests, and focused on the results of the functionalities tested. The analysis involved the following steps.

- Collection of test reports for the campaigns of **¡Error! No se encuentra el origen de la referencia.**.
- Analysis of the case-study test reports provided by the IMs (test specifications and test reports). The idea was to have realistic results by analysing a number of campaigns chosen because they were recent and deemed as representative of current practice. In this early phase, the analysis was not addressed at assessing the systems being tested or the lab/on-site test protocols themselves, rather the viability of the analysis process and its suitability to produce useful information and test accuracy indicators. The analysis involved:
    - Reading the test specifications and classifying each single test into one or more of the defined test categories (functionality by functionality, see **¡Error! No se encuentra el origen de la referencia.**);
    - Identifying and flagging the cases in which an on-site test could be considered as a repetition of an already executed lab tests for the same functionality;
    - Quantifying the identified test accuracy indicators (see subsequent section §2.2.2) classifying each test result into one of four categories;
    - Counting the numbers of tests within the above categories and cases;
    - Analysing and assessing the figures obtained.

Table 1. Past test campaigns used for the study.

	LAB	ON-SITE
ADIF	**80**	**85**
L2 test campaign Madrid-Levante line Albacete RBC. CEDEX-INECO. 2014-2015.	34	27
ERTMS level 2 report, Valladolid-León line. CEDEX 2016	46	0
ERTMS level 2 report, S-100 Equipment Albacete-Alicante	0	58
MULTITEL	**72**	**0**
Campaign 1	30	0
Campaign 2	34	0
Campaign 3	8	0
RFI	**198**	**52**
Design and implementation of the upgrade to standard 2.3.0d. 2016 of the ERTMS line signalling subsystem for the high-speed line Rome-Naples. RFI 2016.	198	52
RENFE	**0**	**84**
Campaign 1	0	84

The categorisation of the ERTMS/ETCS functionalities was inspired following the instructions of TSI 2016/919. For the purpose of this TSI, an 'operational test scenario' means the description of the intended railway system operation in situations relevant for ETCS and GSM-R by means of a sequence of trackside and on-board events related to or influencing the Control-command and Signalling subsystems and the specified timing between them.

The operational tests scenarios are based on the engineering rules adopted for the project.

In some of the test campaign analysed the test reports showed the results already classified by the functionalities below explained and in other cases on test case included several functionalities and these cases, all functionalities tested in each test case have been taken into account.

Table 2 Categorisation of ERTMS/ETCS functionalities.

N.	Category	Description
1	SoM	ATAF procedure: -SoM with invalid or unknown location information. -SoM with valid position -SoM with TSR
3	EoM	Mode transition to SB mode.
10	Level transition	Train approaching a L2 transition border with MA stored and the status of the first signal changes from proceed to non-proceed.
11	Level transition	Level transition although the level transition order or the announcement transition is not received. Verify transitions between (LNTC<=>L2 , L0<=> L2). LNTC to L2 only has transition order, with no announcement.
13	Level transition	Nominal level transitions.
16	Movement authority description	Supervision of the static speed profile with the train running from the beginning to the end of the line at maximum speed.

19	Movement authority description	MA after a SoM with known position.
20	National values	Information regarding National values is sent from trackside at every entry of the L2 area.
23	Obtaining moving authority	Verification of a MA request when the RBC can/cannot extend the Movement authority.
25	On-Sight Protection	Mode transition. Verify the EVC switches modes from FS/SR to OS and that the start location and the length of the OS are is defined according to the infrastructure requirements.
26	On-Sight Protection	Mode transition. Exits from On sight are to FS/SH at a signal. Transition to SH mode ordered by the RBC.
27	Override	Mode transition from SH/ FS to TR when perform a SPAD at an EoA.
28	Override	Override with authorization in FS/PT-OS/TR mode. Verify the "Override" function is available.
30	RBC/RBC Handover	Management of overlapping TSRs in handover area. Supervision of the permitted speed is performed correctly.
32	Rules for balises	The train is running from the beginning to the end of the line at the maximum permitted speed, linking information sent by the RBC complies with balise groups defined in the Track Layout of the line.
33	Shunting	Entering a shunting area -Mode transition from FS/SB to SH ordered by trackside. -Mode transition from FS to OS and after to SH ordered by trackside. -Mode transition from OS to SH. Valid position within the SH area. RBC authorise the entry in SH - If applicable, rejection of the mode change by the RBC for specific situations -Mode transition from SR/PT to SH ordered by the driver.
34	Shunting	Shunting limit overpassed Transition mode from SH to TR
36	Track conditions	Verification that the RBC sends information related the powerless section, the radio hole and the track station platforms condition according to the trackside requirements.
37	Text messages	Verify that a text message is displayed according to the infrastructure rules.
38	Temporary Speed Restriction	Management of TSR information sent by RBC or balise, in FS/OS/SR mode. The EVC manages correctly a revocable TSR.
39	Temporary Speed Restriction	Management of the MRSP when a TSR is established in an adjacent track. Management of the MRSP when several TSR have been established.
40	Temporary Speed Restriction	EVC management of the overlapping of TSR information sent by RBC. FS mode.
41	Temporary Speed Restriction	TSR revocation according to infrastructure rules.
42	Degraded Scenarios	Unconditional emergency stops.
47	Staff Responsible	PT/SR-FS PT/SR
48	Staff Responsible	SR-TR

49	Stand-By Mode	Transition FS/OS/SR-SB mode
50	Supervision	Geographical position.
51	Supervision	Brakes command
52	Supervision	SSP exceeded (Test case 16).

2.2.2 Test accuracy indicators

In the Test Accuracy Framework, the indicators are quantifiable variables that characterise the actual assessment accuracy, for which a single quantity is not sufficient in the current state of the art (whereas e.g. for mechanical measurements, such as a length, a velocity etc. the accuracy is characterised by a single value: the standard uncertainty associated with the measured value).

A-priori indicators

The list in **¡Error! No se encuentra el origen de la referencia.** shows the categorisation of the a priori indicators (IPR) chosen for the VITE project. It was described in Deliverable D2.2 [VITE-WP2-INE-DEL-2 2-v1.0_Test process framework] by limiting to scope to the items that were considered non-negligible for ERTMS/ETCS testing. As an example to further clarify the meaning of the indicators, we consider indicator IPR1 related to ambiguous definitions. This means that the same protocol could be implemented e.g. in two different ways and this could lead to different results (e.g. pass instead of fail). Such a possibility would show up if the test were repeated several times with different operators in different labs. However, this is seldom possible and the possibility remains as a cause of inaccuracy of the test: some time or other the error will occur and the test will give an inaccurate (wrong) result (this would be picked up by an a-posteriori analysis). Sometimes the error will occur but the test result will remain the same. What is counted with this indicator is the number of known occurrences in which such an ambiguous definition was identified and assessed as being a potential cause of a wrong test result, and similarly for the other indicators.

These indicators were created after several internal meetings with on-site and lab expert testers. Different expert judgement have been taking into account during the uncertainty analysis process, in fact, the “A priori” indicators has been modified several times throughout this process.

Table 4 shows the conclusion obtained after the expert testers inputs and the final “A priori” indicators selected for this project.

Table 3 Final “A priori” test accuracy indicators used for the VITE project.

"A priori" indicator	Description
IPR1	n. of known occurrences of ambiguous definitions in test protocols leading to false recordings
IPR2.1	n. of known occurrences of human error (description of equipment in the lab and on site, manufacturers information requested).
IPR2.2	n. of known occurrences of human error during the test execution.
IPR3	n. of known occurrences of error due to the configuration information missing. Complete conditions not available.
IPR4	n. of known occurrences of errors due to simulation impact (delays, accuracy)

Table 4 Checklist for VITE lab tests with observables for "A priori" analysis.

OBSERVABLES DEFINITION	
IPR	**Description**
IPR1	*N. of known occurrences of ambiguous definitions in test protocols leading to false recordings*
Location	The exact location where the test has to be executed is not well defined in the test protocol to make sure that the test is performed in the exact area, at the exact signal...
Test execution conditions	The complete conditions of the test execution are not defined in the test protocol. Number of repetitions needed, the most proper way to execute the test case.
IPR2.1	*N. of known occurrences of human error (description of equipment, manufacturers information requested).*
Equipment versions	The manufacturers have not sent the version of each piece of equipment used during the test execution.
Requirements	The Railway Undertaking has not sent the requirements needed to perform all the test cases.
Manufacturers information	The manufacturers have not sent all the information about the equipment used. They have sent all the information requested by the testers and labs.
IPR2.2	*N. of known occurrences of human error during the test execution.*
Staff misunderstanding	Human error due to the misunderstanding between the testers and the rest of the staff present during the test execution: control centre, signalman, maintenance team… (e.g. wrong KP)
Acknowledge	Acknowledge or do not acknowledge the balise information on time, e.g. while testing the brake command, the tester should not acknowledge in order to check that the brake command is triggered.
Inoperable situations	Even if all the processes are performed correctly, the test case can result in an inoperable situation, e.g. entering a OS or SH area with extremely low speed (speed gradient error).
IPR3	*N. of known occurrences of error due to the configuration information missing. Complete conditions not available.*
Equipment missing	Some devices or piece of equipment are missing (especially in the lab) and the test case cannot be performed.
Track conditions	Information about the track condition is missing. E.g. In the lab only isolated cases are studied, the whole track is not simulated so it could introduce error in some cases.
IPR4	*N. of known occurrences of errors due to simulation impact (delays, accuracy)*
Odometer	Odometer error due to the devices used during the test execution
Interlocking	Significant errors introduced due to the interlocking simulation impact
Simulation delays	Error introduced because of the simulation delays (in the lab)

Equipment disconnected	In order to prevent any issue with the on board equipment, some laboratories disconnect the radio communication while they are executing the test.

The information that should have been collected in the “A priori approach” was not found in the project reports analysed in this project so in order to achieve this information in future campaigns the observable template below, was created. The “A priori” information will be included in this section of the test reports. The results of the “A priori” approach will be analyse in detail in further studies.

Table 5 Observable template

OBSERVABLES				
"A priori indicator"	Observable	Test result	Test case	Comments

A-posteriori indicators

With the analysis and conclusions reached so far, it is considered that these indicators do not correspond to Test accuracy indicators as per the academic definition of test accuracy. However, it is considered that the methodology is useful for 2 main reasons: the conclusions reached with these indicators are considered useful to increase the confidence in the laboratory test results and the indicators already achieved are expected to be able to derive in actual test accuracy indicators if and when there are additional test results inputs to the methodology.

Table 6 shows the a-posteriori indicators considered initially. The analysis of the reports as made available for VITE has shown that it is not currently possible to identify occurrences for IPO2 and IPO3. In fact, normal test reports are obviously not drafted having in mind the purpose of transferring on-site testing to lab testing, and therefore the information to identify false fails or false passes is not available. This is one of the main reasons why it is considered difficult to consider the A-posteriori indicators currently included in the methodology as Test accuracy indicators.

Table 6 A-posteriori test accuracy indicators initially required for the VITE analysis.

"A posteriori" indicator	Description
IPO1	for each functionality, for available lab tests and on-site tests: ratio of ‘fail’ to total number of tests
IPO2	for each functionality, for available lab tests and on-site tests: ratio of false ‘fail’ to total number of tests
IPO3	ratio of false ‘fail’ to total number of tests ratio of false ‘pass’ to total number of tests
NT	total number of tests available for each functionality test category

Therefore, indicator IPO1 was retained and quantified. The problem of not being able to retrieve the information on false passes or fails was partially solved through a sub-class of a-posteriori indicators that was identified within VITE: test history indicators (ITH). Although in these indicators it is not recorded whether the test result (Pass or Fail) is actually true or false, there is some usefulness for test accuracy considerations.

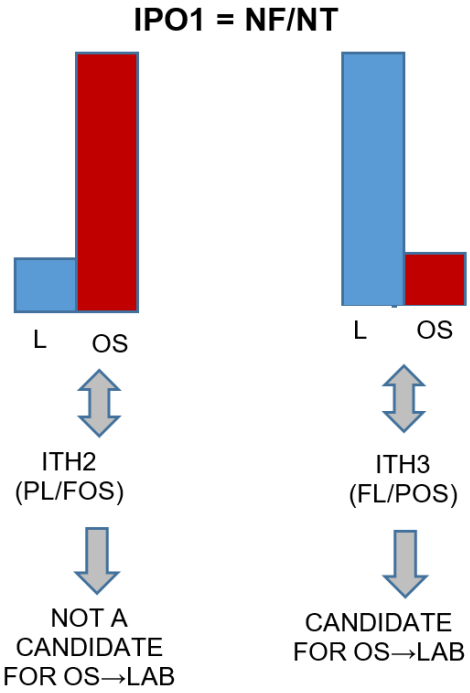
After the analysis of the test campaigns received from RFI, MULTITEL, RENFE and ADIF the “A posteriori” indicators used in this project are shown in the table below, 7.

Table 7 A-posteriori test accuracy indicators currently used for the VITE analysis.

"A posteriori" indicator	Description
IPO5	n. of total ‘fail’ tests
IPO6	n. of total ‘pass’ tests
IPO1	ratio of ‘fail’ to total number of tests
IPO4	ratio of 'pass' to total number of test

IPO5 and IPO6 provides information related to the result itself. IPO1 and IPO4 shows the percentage of pass/fail test in the laboratory and in on site individually, this information does not provide relevant values but the difference between them does, and is related to the test history indicators explained below.

Figure 2 Relationship between IPO1 cases and the test history indicators indicating lab/on-site mismatches.

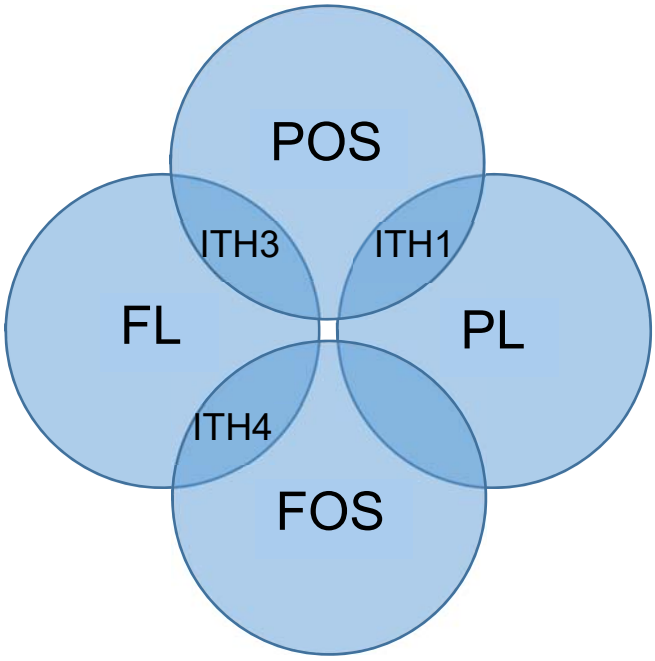


Test history indicators

Test history indicators compare the test results obtained in the laboratory and on site providing the information that constitutes the bases of the shift from on site to lab testing.

In particular ITH2 (PL/FOT) emerged as probably linked to test accuracy issues in spite of the current level of available test report information. In fact, if a test for a given functionality is passed in the lab test, it is less likely that the systems will be changed before the on-site test.

Figure 3 Test history indicators



Test history indicator	Description
ITH1	PL-POS, OK/OK **Pass** in the lab test **Pass** in the on-site test
ITH2	PL-FOS, OK/NOK **Pass** in the lab test **Fail** in the on-site test
ITH3	FL-POS, NOK/OK **Fail** in the lab test **Pass** in the on-site test
ITH4	FL-FOS, NOK/NOK **Fail** in the lab test **Fail** in the on-site test

2.3 Principles

Having described the general methodology, it is possible to relate it more effectively with the theoretical principles of uncertainty analysis, inspired by mechanical measurement standards and transferred to Pass/Fail testing.

First of all, it has to be noticed that the uncertainty it has been addressing means that the actual test result (a measured length, an ERTMS functionality test) is uncertain[1]. Before performing the test (a priori) we are uncertain of the result, and once we have performed the test (a posteriori) we are uncertain on how close it is to a "true value".

Something could have happened during or before the test execution (or measurement) that gave a false result, possibly unnoticed.

If we analyse what could go wrong during or before the test to cause such a false result, we are performing an a-priori analysis - e.g. with the test protocol available, are there any ambiguities that could lead different operators/labs to different results? What is the probability of this occurring? As a consequence, what is the probability of obtaining a false result?

For measurements, the assessment value is a continuous value, so the uncertainty is revealed by performing several (ideally infinite) measurements and observing that there is a variability in the result. The same process is more difficult with Pass/Fail testing, because the variability of the potential causes of an inaccurate (false) result does not always actually cause an inaccurate result. Therefore, generally a large number of repetitions of the test are required in which all the possible

[1] Uncertainty by definition is quantifiable - its qualitative counterpart is inaccuracy, so analysing uncertainty also means analysing (and quantifying) accuracy.

influence variables (causes of inaccuracy) are varied at random before a false result is actually observed. This relates to the concept of reproducibility, as opposed to repeatability.

If it were possible to perform a huge number of reproducibility tests on the same functionality, it would be possible to observe given fractions of Fail and Pass with respect to the total number of tests NT. If e.g. the fraction of F/NT is low and that of P/NT = 1 - F/NT correspondingly high, it is more likely (but not certain) that the "true" result is Pass and that the Fail are due to inaccuracy. In such a case, we would say that the test method is accurate. If the P/NT and F/NT are roughly the same then the method would be judged as highly inaccurate, and it would be impossible to judge with that method whether the "true" result should actually be P or F. This brings us to the two components usually analysed when analysing accuracy: trueness and precision.

Analysing trueness means understanding how close the results are to a "true" value, analysing precision means understanding how reproducible results are. A test method that always gives F on a system that is actually working correctly is highly precise, but completely untrue. For our purposes, similar cases are quite unreasonable, so we can assume that generally the fraction of false F are very low - we are dealing with systems with high trueness. Two different test methods of the same system, one that gives one false result on average every 100 tests and the other that gives one false result on average every 1000 tests have two different levels of precision, the latter being more precise.

It is this variability in precision from functionality to functionality and from lab to on-site that we are looking for in the a-posteriori analysis. In the a priori analysis, we are looking for the causes of these false results.

The information needed actually exists at least in the experience of the test operators, but the actual success of this research lies in what is available in the reports, and could be available if the recommendations of this document are followed.

3 RESULTS OF THE “A PRIORI” APPROACH

3.1 Description of practical implementation

In theory, the practical implementation of the a priori approach aims at identifying all the causes of potential test inaccuracies, quantifying the probability of their occurrence, their correlations and the probability that their occurrence would lead to inaccuracies. In that way, the end-result would be an estimate of the precision of the test process for each of the identified functionalities. The estimate could then be compared with the estimate obtained via the a-posteriori approach. Applied to lab tests and on-site tests for the same functionality and across the different functionalities, this would allow a comparison of the accuracy values and the identification of which functionalities show highly accurate lab tests for example.

In practice, in VITE the scope of the analysis had to be kept manageable by focusing on key types of potential causes of inaccuracy. This was achieved by focusing on the laboratory tests, and consulting experts to identify the key aspects influencing the test result, so as to detail better the indicators outlined in **¡Error! No se encuentra el origen de la referencia.**. This activity resulted in the list of **¡Error! No se encuentra el origen de la referencia.**. Since it was expected right from the start of VITE that there would not be figures already available for such an analysis, it was planned to source the necessary data from the VITE laboratory test campaigns. Thus, the indicators were turned into observables for the tests.

The probability of occurrence for a given a priori indicator and the probability that their occurrence could lead to a false test result are obtainable in theory through the analysis of the observables recorded during the tests. The quantity of test results available for VITE proved not to be sufficient to initiate such a quantification. A continued application of the methodology described in this report would allow the population of the test accuracy framework towards such a quantification.

3.2 Suitability of the proposed a priori indicators

The indicators identified in **¡Error! No se encuentra el origen de la referencia.** were firstly reanalysed by imagining their quantification for a variety of test situations, based on experience within INECO. The result was the following Table 8, where the indicators are associated with typical situations. This analysis confirmed the applicability of the indicators identified.

Furthermore, the use of the identified indicators did not lead to any negative feedback regarding applicability from the VITE demonstration tests, which leads to a good degree of confidence in its future applicability to several different test campaigns.

Table 8 Association of practical test situations with the identified a priori indicators.

Nº	Category	Description	IPR	Example
1	Equipment versions	Information about the version of the equipment has not been sent.	IPR2.1	Information about the equipment versions missing.
2	Complete real conditions	No info about lab environment in the test reports. In the lab, the testers simulate just the piece of track they are going to test but not the whole track. For this reason, some error can be introduced since the information of the rest of the track (adjacent tracks) is missing in the laboratory. In the laboratory isolated cases are studied.	IPR3	Complete condition missing while executing test in the laboratory
3	Odometer accuracy	It is simulated in the lab but each on-board equipment has its own odometer module (manufacturer).	IPR4	Odometer error.
4	Interlocking	The interlocking is also simulated in the lab (possible error).	IPR4	error due to the interlocking simulation
5	TSR	KP discrepancies, it would be easier to test these cases in the lab (accuracy).	IPR2.2	PK information more accurate in the lab than on site. The tester introduces the data.
6	Equipment	Some pieces of equipment are missing in the lab so some test cases cannot be performed in the laboratory.	IPR3	Equipment missing in the lab
7	Human error	It is worse on site since in the lab the tester enters the data. (More accurate in the lab than on site).	IPR2.2	Entering in OS area.
		Acknowledge or no acknowledge of balises on time.	IPR2.2	
8	Protocols	Ambiguous definitions in test protocols. The results can be worse on site since if the test is not performed in the proper way maybe you can not repeat it.	IPR1	A light signal test was executed. The test was performed on a signal with no yellow light and that colour was necessary so the test could not be executed.

9	Communication	Track occupation. When you are testing on site sometimes external assistance is needed (e.g. maintenance team). If this case is tested in the laboratory the testers can do it on their own.	IPR2.2	Misunderstanding between the staff involved in the test performances
		In order to prevent any issue with the on-board equipment, some laboratories disconnect the radio communication while they are executing the test.	IPR4	Equipment disconnected
10	Communication	False "pass" because the RU didn´t send all the requirements.	IPR2.1	RU didn´t send all the requirements.
11	Interpret data	Even all the processes are performed correctly, the result can be an inoperable situation. The on-site test is an advantage in this case since when you are testing on site, in the train you can clearly see and detect an inoperable situation.	IPR2.2	Entering in a OS or SH area with extremely low speed (speed gradient error)
12	Manufacturer information Protocols definition	Some of the cases tested on the lab are considered "false pass" since everything is apparently OK but it is actually not because some information is missing. The manufacturer does not provide the information requested. The following aspects should be improved: -Real equipment RBC, EVC. -Interfaces as automated as possible. -TSR manager environment similar to the real ones.	IPR2.1	Manufacturer information missing
13	Delays simulation	All the delays simulated on the lab can introduce some error that should be considered.	IPR4	Lab equipment
14	System capacity	Test considered "false pass" because the system capacity was not tested.	IPR1	Large number of TSR on the same track.

4 RESULTS OF THE “A POSTERIORI” APPROACH”

4.1 Description of practical implementation

In theory, the practical implementation of the a-posteriori approach aims at estimating the precision of the test process for each of the identified functionalities, as means of comparison with that obtained via the a priori approach. Applied to lab tests and on-site tests for the same functionality and across the different functionalities, this would allow a comparison of the accuracy values and the identification of which functionalities show highly accurate lab tests for example.

In practice this goes through the quantification of the a-posteriori indicators for which the necessary information is available, and the formulation of recommendations on what information the reports should provide in order to arrive at increasingly improved uncertainty analyses.

The core of the results used for the a-posteriori analysis is formed by the ADIF and RFI test campaigns. The key issue to be addressed was whether the test reports from different parties could be analysed effectively with the proposed methodology.

Some of the test repots received, contained results that were already categorised in a similar way as per the proposed functionality test-category classification. The test descriptions were brought together in a test-case worksheet (**¡Error! No se encuentra el origen de la referencia.**), and then a counting process (**¡Error! No se encuentra el origen de la referencia.**) led to the quantification of the indicators (**¡Error! No se encuentra el origen de la referencia.**).

Table 9 Excerpt of test case worksheet results.

TEST	TEST DESCRIPTION	LAB	ON SITE	MODE	LEVEL	AIM	TEST PERFORMANCE	RESULT REASONS
1	2.1.2 Static speed profile supervision.	OK	OK	FS	2	The aim of this test case is to verify the system intervention when the permitted speed of the SSP is exceeded.		
2	2.3.1.1 MA update in FS mode. New EoA at a light signal. Train in movement.	OK	OK	FS	2	The aim of this test case is to verify that the EoA is updated when the train receives a shortened MA (with new EoA in a light signal) followed by an extension of the MA.		
3	2.3.1.1 MA update in FS mode. New EoA at a light signal. Train at standstill in front of the signal.	NO OK	OK	FS	2	The aim of this test case is to verify that the EoA is updated when the train receives a shortened MA (with new EoA in a light signal) followed by an extension of the MA.	Light signal E1 Tarazona is set in proceed aspect. Train stops in front of the light signal E1 and it is requested to the signalman to switch the signal E1 to stop aspect. The MA should have been updated , however message 16 (Unconditional Emergency Stop) is sent to the RBC and the train switched to TR.	NO OK because, if the train is stopped in front of a light signal E1 in stop aspect , RBC sends a unconditional emergency stop after sending message ''shorten MA".

Table 10 Indicator count for functionality test category n. 28 'Override with authorization in FS/PT-OS/TR mode. Verify the "Override" function is available.'

functionality test category n. 28 'Override with authorization in FS/PT-OS/TR mode. Verify the "Override" function is available.'			
N	Description	LAB	ON-SITE
16	2.13.1.1 Perform a SPAD at an EoA. Mode transition from FS to TR at a stop signal.	OK	
18	2.13.2.1 Perform a SPAD at an EoA. Mode transition from OS to TR at a stop light signal.	OK	NO OK
23	2.14.1.1 Override with authorization. FS mode . The radio communication session is established with the RBC.	OK	OK
24	2.14.2.1 Override with authorization. OS mode . The radio communication session is established with the RBC.	OK	
66	2.13.2.1 Perform a SPAD at an EoA. Mode transition from OS to TR at a stop light signal.	OK	
69	2.14.1.1 Override with authorization. FS mode . The radio communication session is established with the RBC.	OK	
70	2.14.5.1 Override with authorization. PT mode. The radio communication session is established with the RBC.	NO OK	
125	2.14.5.1 Override in PT mode. Communication with RBC active.		NO OK
	NT	7	3
	NF	1	2
	NP	6	1
	ITH1	1	
	ITH2	2	
	ITH3	1	
	ITH4	1	

For the above functionality, there are a few indicators that provide information on how lab tests perform versus on-site tests. Following the methodology previously explained, the values of the indicators would be as follows:

- fraction NF/NT lab vs on-site (1/7 vs 2/3)
- ITH1 = 1
- ITH2 = 2
- ITH3 = 1
- ITH4 = 1

At the beginning of the project, another way to analyse the results of the test campaigns was considered, following this method only the test that were executed in the laboratory and repeated on site, were taking into account and therefore in the example shown in table 10, the values of the test history indicators would have been:

- ITH1 = 1
- ITH2 = 1
- ITH3 = 0

- ITH4 = 0

Following this method test n. 18 led to the ITH2 count. However, the low amount of data available (i.e. for ERTMS it is rare to perform the exact tests on site and in the laboratory) and the concerns that the exact same test was performed on site and in the laboratory (e.g. slight modifications of the products or the engineering) derived in the selection of the first criteria described for the methodology.

Therefore, even of the indicators included in the methodology are considered not compliant with the academic definition of the test accuracy, they are considered useful indicators to increase the confidence of the results performed in laboratory based on the different ERTMS functionalities.

In **¡Error! No se encuentra el origen de la referencia.** the uncertainty indicator information is summarised functionality by functionality.

Table 11 Excerpt of uncertainty indicator

FUNCTIONALITY TEST CATEGORY			UNCERTAINTY INDICATORS																				FAILED-TEST HISTORY INDICATORS				Test case
N.	Category	Description	IPR1		IPR2		IPR3		IPR4		IPR5		IPR6		IPO5		IPO6		IPO1		IPO4		ITH 1	ITH 2	ITH 3	ITH 4	Test number
			Lab	On-site	Lab	On-site	Lab	On-site	Lab	On-site	Lab	On-site	Lab	On-site	Lab	On-site	Lab	On-site	Lab	On-site	Lab	On-site					
23	Obtaining moving authority	Verification of a MA request when the RBC can/cannot extend the Movement authority.	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	100%	0%	0%	100%	0	0	1	0	Test 33
25	On-Sight Protection	Mode transition. Verify the EVC switches modes from FS/SR to OS and that the start location and the length of the OS are is defined according to the infrastructure requirements.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	3	0%	0%	100%	100%	3	0	0	0	Tests 9, 10, 55, 56, 57, 95
26	On-Sight Protection	Mode transition. Exits from On sight are to FS/SH at a signal. Transition to SH mode ordered by the RBC.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	0%	0%	100%	100%	2	0	0	0	Tests 7, 49, 85
27	Override	Mode transition from SH/ FS to TR when perform a SPAD at an EoA.	0	0	0	0	0	0	0	0	0	0	0	0	1	1	2	1	33%	50%	67%	50%	1	1	1	1	Tests 17, 20, 68
28	Override	Override with authorization in FS/PT-OS/TR mode. Verify the "Override" function is available.	0	0	0	0	0	0	0	0	0	0	0	0	1	2	6	1	14%	67%	86%	33%	1	2	1	1	Tests 18, 23, 16, 24, 66, 69, 70, 125
30	RBC/RBC Handover	Management of overlapping TSRs in handover area. Supervision of the permitted speed is performed correctly.	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0%	100%	0%	0%	0	0	0	0	Test 137
32	Rules for balises	The train is running from the beginning to the end of the line at the maximum permitted speed, linking information sent by the RBC complies with balise groups defined in the Track Layout of the line.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0%	0%	0%	100%	0	0	0	0	Tests 102, 103, 104, 105, 106

Other test cases were reported in a different way since several functionalities were examined simultaneously. The analysis of the test description, (similar worksheet as that of **¡Error! No se encuentra el origen de la referencia.**) was followed by a count of how many times each functionality was tested per single test.

The process developed thus provides the possibility to merge the results of test reports from different parties simply by summing e.g the number of ITH3 cases etc., functionality by functionality. This is a key finding as regards the applicability of the proposed process at the EU level.

4.2 Indications on minimum number of tests required

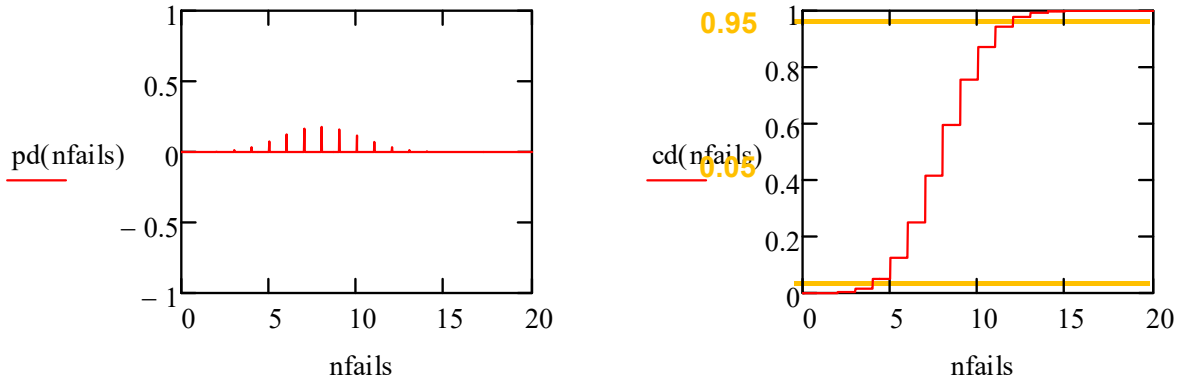
A question that arose during the research was the minimum number of tests required for the results to lead to significant conclusions. The following statistical considerations provided an answer to this. It was not possible to identify alternative a-posteriori indicators that could incorporate these considerations in one single figure this is a possible development of the research.

Let us consider as the random variable of interest the number n_{Fs} of tests with result 'Fail' in a sample of N tests on the same functionality. The ratio $\theta_s = n_{Fs} / N$ indicates an estimate of the fraction of failed tests of the entire population of tests. The estimate is affected by uncertainty, since if we repeated N tests on the same object the n_F would change, and if we repeated it an infinite number of times it would draw out a binomial distribution, much like that associated to tossing a coin N times. However, whereas for the coin the assumed fraction of 'heads' or 'tails' is known, 0.5, for our problem the fraction θ is unknown and estimated through the sample of N tests.

In order to quantify the uncertainty associated with the sample estimate, we look for the Lower and Upper Confidence Limits for variable N and n_F. That is to say, if e.g. we have N = 20 tests and we feel that about 4 out of 10 are expected to fail (e.g. because we have had 8 fails, or for other reasons), we calculate the value of n_F that will not be exceeded by only α = 5% of a large number of samples (LCL) and the value of n_F that will not be exceeded by α = 95% of samples (UCL). These turn out to be 4 and 12 respectively (see **¡Error! No se encuentra el origen de la referencia.**). This means that we are highly confident that the actual fraction θ lies between 4 / 20 = 0.2 and 12 / 20 = 0.6. In other words, the estimate of nF = 8 is affected by an uncertainty of (0.2-0.4)/0.4 to (0.6-0.4)/0.4, i.e. (-50%, +50%).

The above applies under the assumption of tests for the same functionality within a given campaign by a given party that are independent. This is probably not the case for such tests. As for the TrioTRAIN project, given the complexity of the issue, the assumption is held as true in a first step, and the results are checked to ensure that it does not have a significant influence on the conclusions.

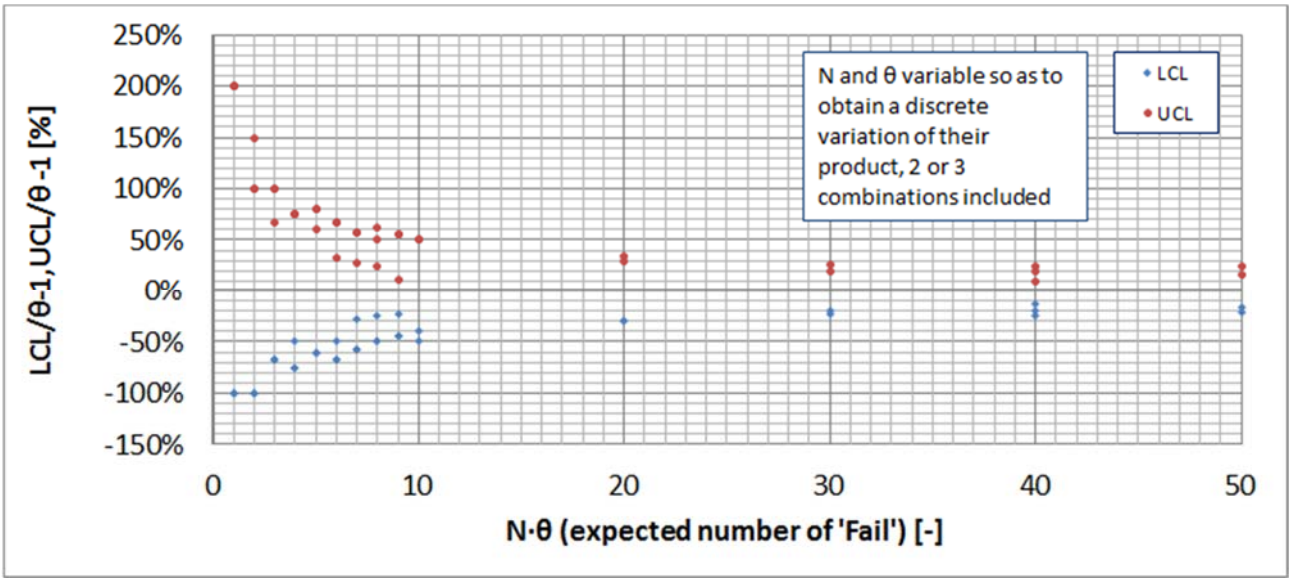
Figure 4 Probability density (PDF) and cumulative distribution (CDF) functions for the binomial distribution (N = 20, θ = 0.4, N·θ=8). The 95% confidence interval in terms of nF (4, 12) or θs = nF/N (0.2,0.6) or θs/θ (-50%,+50%) is shown on the CDF.



In order to establish a minimum number of tests required to have a reasonably good sample, we now examine how the confidence limits vary with the number of available tests N and the fraction of 'Fail' θ. Actually, **¡Error! No se encuentra el origen de la referencia.** shows that it is sufficient to examine the trend (of the relative uncertainty on the actual fraction of 'Fail') as a function of the product $N{\cdot}\theta$, which is the expected number of 'Fail'. In fact, in order to construct the chart, both N and θ very varied within large ranges e.g. N between 10 and 1000 and θ in a way as to obtain integer values for the product as seen on the x-axis. It can be seen that the result depends more on the product $N{\cdot}\theta$ than on the single values of the two factors. In other words, in order to obtain a sample in which the fraction nF / N is representative to within ±50% the product $N{\cdot}\theta$ has to exceed approx. 10, meaning 10 'Fail' almost independently of whether the come from a sample of 100 (θ_s = 0.1) or of 50 (θ_s = 0.1). It can also be observed that there is only a slow improvement for higher values of the product, with a uncertainty of ±10% achieved only for $N{\cdot}\theta \geq 250$. **In practical terms the criterion**

$N \cdot \theta \geq 10$, (estimated through $N \cdot \theta_s = n_{Fs}$ number of 'Fail' in the sample) appears to be a good starting point.

Figure 5 Variation of upper and lower confidence limits with N·θ (expected number of failed tests).



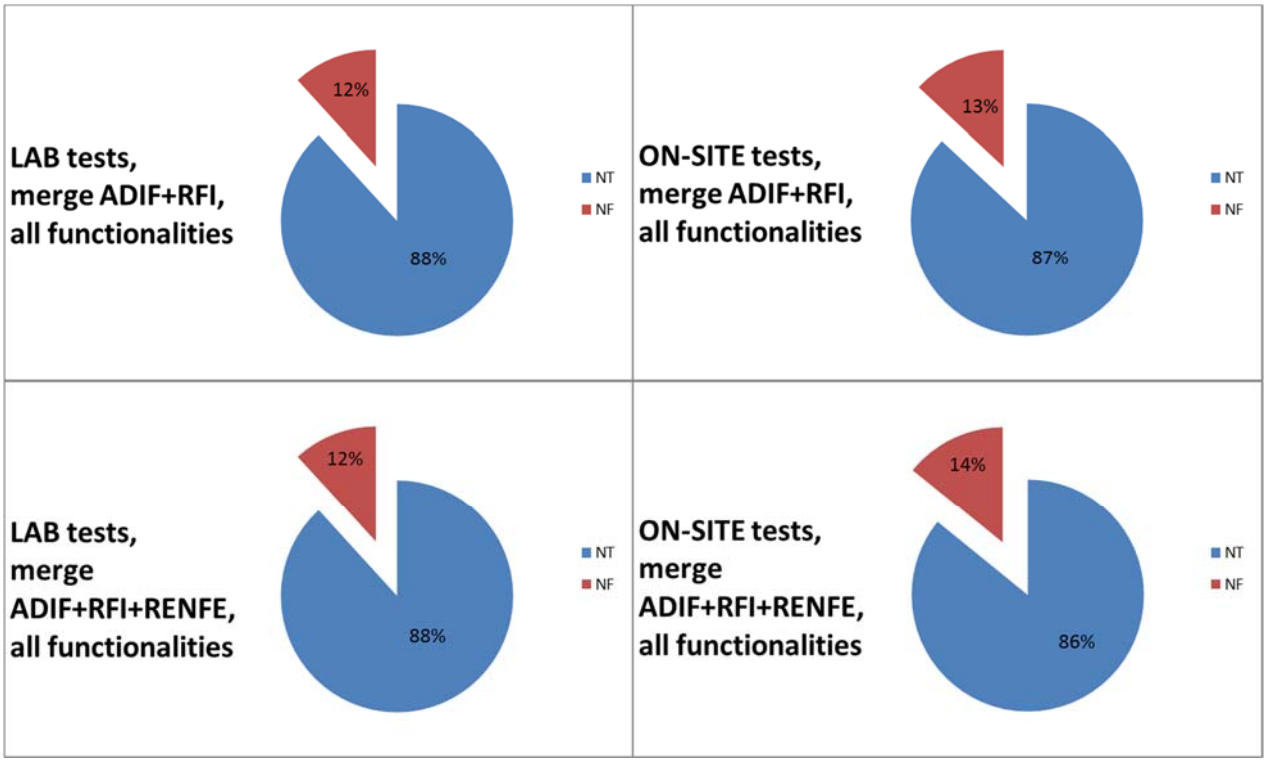
4.3 Quantification of the key a posteriori indicators based on the VITE test campaigns

The available test campaigns for the quantification of the a-posteriori indicators and the amount of tests for each, are shown in **¡Error! No se encuentra el origen de la referencia.**. RENFE contributed only on-site tests whereas Multitel contributed only lab tests.

Figure 6 Number of tests available according to the contributor and classified as lab / on-site.

In **¡Error! No se encuentra el origen de la referencia.** the overall results are analysed by highlighting the proportion of fails with respect to the total number of tests. It is interesting to note that both, lab tests and on-site tests show a fraction of failed tests of 10% to 15%. This changes when the analysis goes into the detail of the functionalities. The charts show the situation with and without the RENFE tests, as a measure to provide an idea of the sensitivity of the figures to considering or not a certain campaign (with only on-site tests in this particular case). There is not much change in the proportions identified above.

Figure 7 Fraction of failed tests in the test campaigns, lab vs on-site.



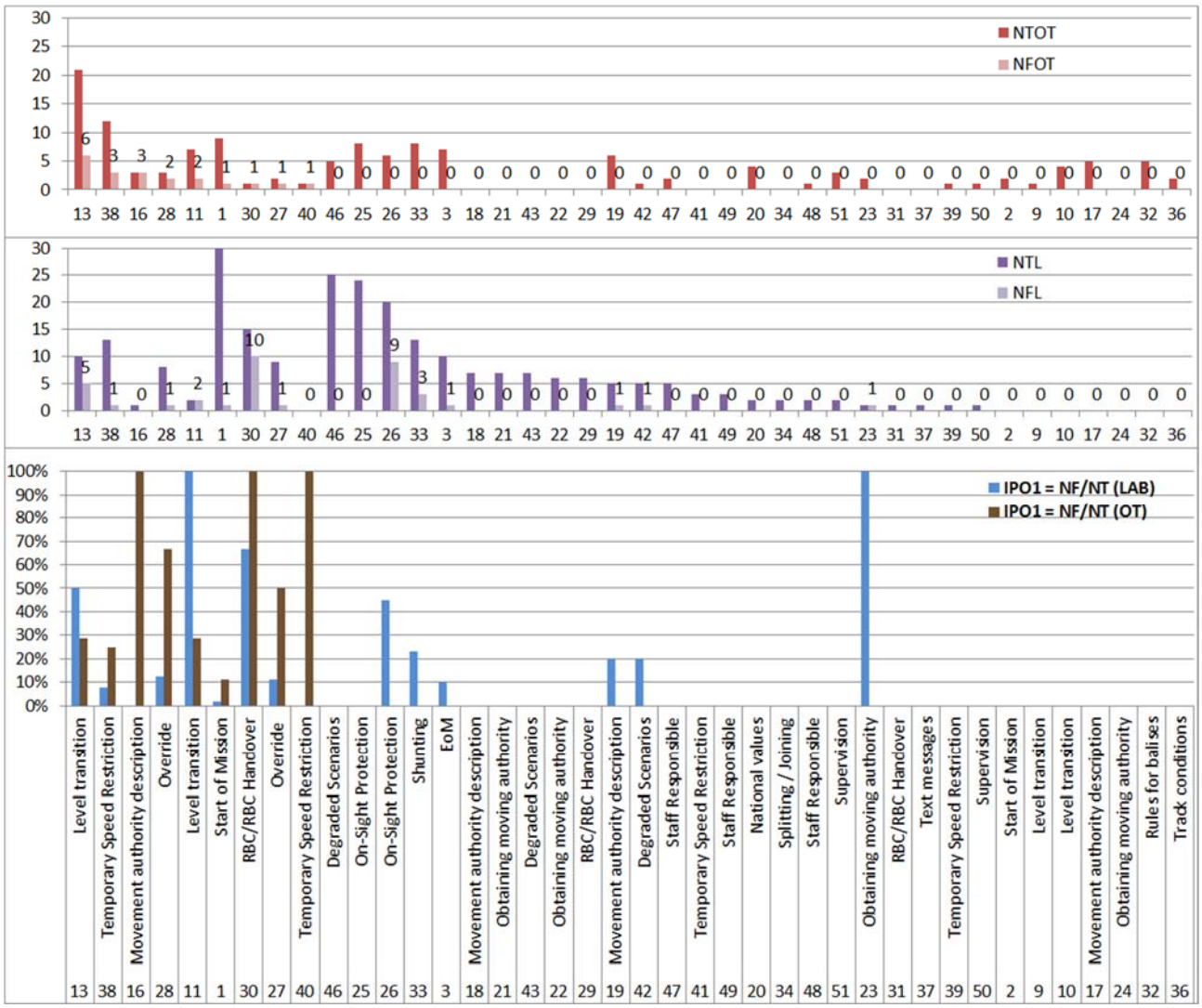
The analysis of the ADIF+RFI merged results by functionality is shown in **¡Error! No se encuentra el origen de la referencia.**, where the aim is the quantification of a posteriori indicator IPO1 (NF/NT) for both Lab (L) and On-site (OS) tests. In the upper and middle charts the figures for NF are shown in order to show compliance or not with the minimum-number-of-tests criterion given in §4.2. It may be seen that the test campaigns available are not yet sufficient to provide good IPO1 quantifications. Again, the methodological aspects behind the creation of the charts are the ones that are most interesting for the purposes of this report.

During the analysis, it was observed that it is interesting to draw up the charts with different sorting criteria as they highlight different aspects. In **¡Error! No se encuentra el origen de la referencia.** the results are sorted by decreasing number of failed on-site tests. This, highlights the compliance with the minimum-number-of-tests criterion, which unfortunately is failed here. More campaigns would be needed for a large number of functionalities to have at least 10 failed tests observed. A number of campaigns of the order of 10 times that available for VITE would lead to the first 9 functionality test-categories (up to category 40) complying with the criterion for the on-site tests, assuming that the NF/NT ratios remain stable. Similarly, for the lab tests 13 functionalities would have meaningful results if 10 times the test results were available.

Observation of the bottom histogram is the instrument to start assessing the suitability of the lab tests versus on-site tests for each functionality. If we imagine that the upper and middle chart showed values compliant with the minimum-number-of-tests criterion, we could take the IPO1 fractions in the lower chart to be reliable and look for the cases shown in **¡Error! No se encuentra el origen de la referencia.**.

Figure 8 Analysis per functionality of the merged ADIF+RFI results: total numbers of tests per functionality, On-site & Lab (NTOS, NTL), corresponding numbers of Fails (NFOS, NFL),

and a posteriori indicator IPO1 (NF/NT, Lab and On-Site). Results sorted by 1\\\decreasing number of on-site fails.



When a given functionality, over a large number of tests, has shown a low value of lab fails but a high value of on-site fails, and regardless of the additional information provided in the reports, it could be an indication the on-site tests are required. This indication would have to be analysed in conjunction with that of the ITH2 indicator (PL/FOS) to which it is in a way related. In fact, if in the reports we are able to count the number of times that a passed lab test was followed by a failed on-site test for the same identical system (ITH2), this indicator should also be high to confirm the IPO1 indication.

5 CONCLUSIONS

This paper has defined a framework and a methodology for:

- the analysis of existing test campaign reports / results ("a posteriori analysis") to highlight the information that can be used to indicate the areas in which lab testing can be considered as a candidate for partial or complete replacement of field tests;

- the analysis of past tests and VITE demonstration tests to quantify the indicators that can provide information on the (potential) causes of laboratory and on-site test inaccuracy ("a priori analysis"), in support of the a-posteriori analysis.

For the a-posteriori analysis, the developed ERTMS/ETCS test accuracy framework was populated with the results from realistic case-studies, based on actual test reports provided by the VITE consortium. Past test results were analysed for the a priori analysis, and a list of observables was developed for the VITE demonstration tests which provided preliminary results as to the suitability of identifying the causes of laboratory inaccuracies while performing the tests.

The main findings are summarised below.

5.1.1 On-site to lab shift

- As expected, the available test results were not sufficient at the moment to draw robust conclusions as to which functionalities are candidates for such a shift.
- However, the methodological indications given in this deliverable are important to populate the framework and arrive at more meaningful quantifications in support of on-site to lab shifts on an ERTMS/ETCS functionality basis.

5.1.2 Roadmap for populating the Test Accuracy Framework in support of the on-site to lab shift

- The actors that could be involved in populating the framework are the suppliers of ETCS systems, the European Agency for Railways, Notified Bodies, Infrastructure Managers and Railway Undertakings. The idea is that the test reports already required when placing such systems in service contain quantitative information that can support the on-site → lab shift. However, such reports are not drafted for this purpose, so important improvements could be made if the information recommended in this deliverable were always included in the reports. The analysis of the data (i.e. an analysis similar to that described in this deliverable) would however have to be done an EU-wide level, and this calls for ERA and EC action.
- The key finding in this sense is that it was possible to merge the test indicators for different campaigns as shown in the report, thus demonstrating, at least for the a posteriori indicators, the possibility to further populate the developed ERTMS/ETCS test accuracy framework at an EU-wide level.
- The charts and tables contained in §3 (a priori analysis) and §4 (a posteriori analysis) of this report may be used as templates for further population of the framework. The charts could remain the same, with the indicator values becoming more and more reliable as the number of available tests increases.
- As the test numbers increase, a comparison of a priori and a-posteriori indicators would provide more insight on the causes of the lab inaccuracies that were only preliminarily identified in this report. In this way mitigation measures on the identified causes could be put in place to further increase the accuracy of the lab tests that are identified as "just-about-mature" to a level that can allow a partial or total on-site → lab shift.