Journal of the Brazilian Computer Society, 2026, 32:1, doi: 10.5753/jbcs.2026.5473

A Guide-based Approach for IoT Interoperability Testing

Karina da Silva Castelo Branco [**Federal University of Ceará** | ***karinascbranco@alu.ufc.br***]
Valéria Lelli Leitão Dantas [**Federal University of Ceará** | ***valerialelli@ufc.br***]
Lavínia Matoso Freitas [**Federal University of Ceará** | ***laviniamatosof@alu.ufc.br***]
Liana Mara Carvalho [**Federal University of Ceará** | ***liana@alu.ufc.br***]
Paulo Ricardo Fernandes Rodrigues [**Federal University of Ceará** | ***paulofr@alu.ufc.br***]

Federal University of Ceará, Av. Mister Hull, s/n, Pici, Fortaleza, CE, 60455-760, Brazil.

Received: 24 January 2025 • **Accepted:** 20 September 2025 • **Published:** 29 June 2026

Abstract

Technology has significantly transformed human interactions with everyday objects, expanding their communication capabilities. This broad connectivity to the Internet has given rise to the "Internet of Things (IoT)", extending the boundaries of the Internet and evolving those objects into smart objects. When interconnected, these smart objects can collect and share information to provide services effectively. However, the intense data traffic and the diversity of interaction methods among these objects bring challenges related to interoperability. Interoperability refers to the ability of distinct systems to communicate effectively, ensuring data integrity. In this context, interoperability testing assesses the capability of systems and devices to cooperate efficiently. Regarding the testing challenges faced in IoT, we can highlight (i) the complexity of the IoT architecture, which involves multiple layers and system components; (ii) the diversity of IoT protocols; and (iii) the heterogeneity of devices, which poses challenges in testing their compatibility and managing their integration to ensure effective connectivity. To address these challenges, this paper proposes a guide-based approach designed to systematize IoT interoperability testing. The guide was developed following a systematic methodology, including literature reviews, data extraction and analysis, and observations in real IoT environments. The guide covers 12 topics, such as the definition and correlation of IoT characteristics, subcharacteristics, abstract test cases, measures, cost-benefit analysis, and tool suggestions. The guide evaluation was conducted in three stages: (1) a structural evaluation using the Technology Acceptance Model (TAM); (2) a controlled experiment applying the guide to test the interoperability of a real IoT application; and (3) an evaluation of the guide with domain experts. The results of the three evaluations were positive, indicating that the guide is effective in test planning, execution, and identifying IoT-specific failures.

Keywords: Internet of Things, Guidelines, Interoperability Testing

1 Introduction

Technology has significantly transformed how humans interact with everyday objects (*e.g.,* refrigerators, watches, air conditioners, and cars [Mattern and Floerkemeier, 2010]), expanding their communication capabilities. This broad connectivity to the Internet has led to the emergence of the "Internet of Things (IoT)", which expanded the Internet's boundaries to include these objects, commonly referred to as "smart". When interconnected, such objects can collect and share information to provide services effectively.

According to Giusto et al. [Giusto *et al.*, 2010], IoT represents a communication paradigm among smart objects, which share information among themselves and other entities over the network to provide services. Thus, IoT enables a wide range of interactions between humans and objects (*i.e.* human-thing interactions) and among the objects themselves (*i.e.* thing-thing interactions), creating an interconnected ecosystem.

The expansion of connected devices has presented various challenges in the development cycle of IoT systems, especially in the testing phase. These solutions consist of heterogeneous devices with distinct communication and interaction characteristics, leading to the development of new protocols and services for IoT systems [Zaidi *et al.*, 2009]. Consequently, these systems rely on infrastructures consisting of heterogeneous networks that require interconnection, *Interoperability*, performance, and security to ensure the desired quality of service [dos Santos, 2016]. Because of this complexity, software testing plays a critical role in that cycle, as it aims to validate and verify the product's functionality, performance, security, and compliance, ensuring it delivers value and meets user requirements [Myers *et al.*, 2011; Coutinho *et al.*, 2016].

Thus, the key testing challenges faced in IoT are related to Security, *Interoperability*, and Performance characteristics [Carvalho *et al.*, 2022]. These characteristics stand out as central for ensuring the efficient and secure operation of IoT applications.

Interoperability, the focus of our study, refers to the ability of two or more systems to communicate effectively, ensuring data integrity and operational efficiency. In this context, interoperability testing verifies the systems' ability to interact consistently and cohesively. This test also evaluates the systems' efficiency of communication and information sharing, ensuring that resources can be accessed and used appropriately across different systems and organizations [ISO/IEC 30141, 2018; Mattiello-Francisco, 2009].

The importance of addressing *Interoperability* issues is frequently discussed in the literature [Soares, 2009; Gulla *et al.*, 2006], particularly because of the *intense data traffic* and *diverse interaction methods* among smart objects.

Interoperability issues arise, for instance, in common IoT scenarios such as smart homes with various devices (*e.g.,* voice assistants, security cameras, thermostats, smart bulbs, and locks), where seamless communication between them is crucial. This scenario introduces challenges related to architectural complexity, which involves multiple layers and system components; the diversity of IoT protocols (*e.g.,* MQTT, HTTP, and CoAP) and standards (*e.g.,* Bluetooth and Zigbee); and the heterogeneity of devices, which poses challenges in testing their compatibility and managing their integration to ensure effective connectivity. Moreover, network connectivity also plays a critical role, as interoperability testing must evaluate the devices' ability to maintain effective communication under different network conditions.

Considering the identified challenges and gaps, this paper presents a guide-based testing approach focused on the *Interoperability* characteristic in the domain of IoT, leveraging the methodology defined in previous studies [Carvalho *et al.*, 2022] [Branco *et al.*, 2024].

To achieve this goal, we investigated the following key research questions (RQ):

RQ1. What aspects should be considered to evaluate the interoperability characteristic in IoT applications?

RQ2. How effective is the guide-based approach in evaluating the interoperability characteristic in IoT applications?

Our approach provides systematized interoperability testing based on the study of [Carvalho *et al.*, 2022]. We have used this study to select interoperability as a challenging characteristic in IoT testing and to leverage our previous experience to build the IoT interoperability testing guide [Branco *et al.*, 2024]. Thus, initially, we conducted a literature review based on the guidelines of systematic reviews proposed by Kitchenham [Kitchenham *et al.*, 2009]. Subsequently, we identified the subcharacteristics of interoperability to detail and assess interoperability more precisely, following the ISO/IEC 25010 [ISO/IEC 25010, 2011]. We also searched for standards, approaches, and tools used in the *Interoperability* testing. The guide covers 12 topics, including definitions and correlations, abstract test cases, measures, and tools to support the Interoperability testing. Also, we provide a new topic in the guide titled "Interoperability Testing Challenges".

To evaluate the guide-based testing approach, we conducted three evaluations: (i) guide evaluation using the Technology Acceptance Model (TAM) [Davis, 1989]; (ii) a controlled experiment [Wohlin *et al.*, 2012] to evaluate the use of the guide for testing the *Interoperability* of an IoT application; and (iii) evaluation with domain experts.

The results of the TAM-based evaluation indicated participants' high acceptance of the guide, highlighting the positive perception of its usefulness and ease of use. The controlled experiment revealed that the guide significantly facilitated the identification of *Interoperability* issues and enhanced the tests' effectiveness, demonstrating its practical effectiveness and applicability to different IoT testing scenarios. The evaluation with experts shows positive results for planning and executing *Interoperability* tests for researchers and industry professionals.

The main contributions of this paper are:

- the guide for IoT interoperability testing, structured in 12 topics;
- the identification and categorization of Interoperability Testing Challenges in IoT;
- the Wiki to facilitate the use of the guide; and
- a set of lessons learned related to developing and validating the guide-based approach.

The remainder of this paper is organized as follows: Section 2 provides the background on IoT Interoperability Testing. Section 3 describes the research methodology. Section 4 presents the testing guide covering its 12 topics for IoT. Section 5 presents the Wiki. Sections 6, 7, and 8 present the three evaluations conducted to assess the guide-based approach. Section 9 discusses the results through the research questions. Section 10 presents the lessons learned. Section 11 presents the threats to validity. Section 12 presents the related work, and finally, Section 13 presents the final considerations and future directions.

2 Background

In this section, we present the central concept of this research: Interoperability testing in IoT.

According to ISO/IEC 15926 [ISO 15926, 2011], Interoperability refers to the ability of different systems, devices, or applications to work together effectively, enabling accurate and efficient communication, program execution, and data transfer.

This capability aims to integrate and enable independent entities to work together, even when they are planned and implemented separately [Motta *et al.*, 2017]. From the user's perspective, Interoperability refers to the ability to access heterogeneous informational resources distributed across multiple network nodes through a single interface without understanding the underlying storage mechanisms [Weiss *et al.*, 2019; Gulla *et al.*, 2006].

Grounded in this principle, Interoperability testing evaluates the ability of different systems, components, or devices to communicate and work effectively and integrally in a specific environment [Andrade and Luque, 2022]. This process evaluates the system's ability to exchange and process information reliably and efficiently, ensuring seamless interaction and collaboration among all components to fulfill their intended functions.

According to ISO/IEC 21823:2022 [ISO/IEC 21823, 2022], Interoperability in IoT environments can be defined as the ability of IoT devices and systems to communicate and cooperate effectively within an integrated ecosystem. This interoperability ensures that different devices, even those from different manufacturers or using varying communication protocols, can interact effectively, enabling efficient data exchange and automated process execution.

Moreover, adopting a structured guide-based approach offers significant advantages for Interoperability testing in IoT, providing systematic, reusable guidelines that help address the inherent complexity of heterogeneous IoT environments. Unlike more ad hoc or layer-specific techniques, guides offer a comprehensive and organized perspective. This promotes standardization of testing activities and clear documentation of procedures, facilitating implementation across different contexts by various professionals, including those who are not software testing experts. This characteristic is particularly relevant given IoT systems' diversity of devices, protocols, and standards.

An effective evaluation of interoperability in IoT environments requires an understanding of the operational context of each device and system in the network, as well as their specific interactions and functionalities. This evaluation requires a detailed analysis of the communication infrastructure and interfaces of each IoT application to ensure that data exchange and action coordination between devices occur efficiently and integrally [Rowland *et al.*, 2015].

The main challenges in Interoperability testing include data heterogeneity, various communication protocols, data transport methods, and access methods, which can render communication inoperable [Nunes, 2011]. These factors pose significant difficulties in converting heterogeneous data into compatible formats, highlighting the need for system and protocol standardization [Andrade and Luque, 2022].

In the context of IoT, despite the abundance of platforms, there is a lack of scientifically grounded guidelines for their use and validation [Cañas Betancur and Hernández Sánchez, 2019].

Methods for evaluating IoT applications include checklist-based inspection techniques. These techniques assess the user experience and are structured into five categories: context, content, structure, images, and usability [Maia *et al.*, 2012; Caldas, 2023; Almeida, 2018].

Conversely, there are testing approaches focused on evaluating critical IoT characteristics. The work [Carvalho *et al.*, 2022] proposes a testing guide for the performance characteristic based on 11 key topics. These topics are derived from the performance sub-characteristics established in [ISO/IEC 25010, 2011], and the guide can be instantiated for other IoT characteristics.

3 Methodology for Instantiating the Guide

We leveraged the methodology based on IoT characteristics defined in a previous work [Carvalho *et al.*, 2022] to develop a testing guide focused on the *Interoperability* characteristic in IoT. The guide's structure is organized into 12 topics, forming the foundation for its instantiation.

The methodology for instantiating the guide structure consisted of six main activities, as illustrated in **Figure 1**.

The activities are presented in the following subsections.

3.1 Literature Review

The first activity consists of "Literature Review". We followed the guidelines based on systematic reviews proposed by [Kitchenham *et al.*, 2009]. We defined the goals, the research questions, and a search string. To facilitate the process, we used the Parsifal[1] tool to plan, conduct, and report the research. The search string focused on the selected IoT characteristic *Interoperability* as illustrated in **Table 1**.

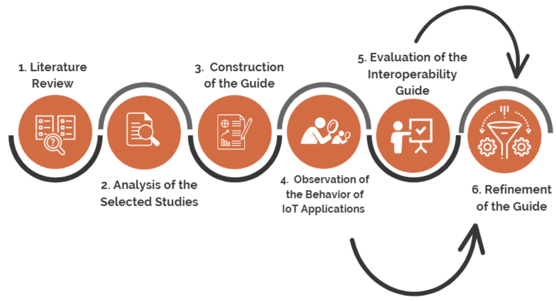

Figure 1. Methodology used to instantiate the guide [Branco *et al.*, 2024].

Table 1. Search String.

("internet of things") AND ("interoperability test" or ("interoperability testing") AND (guideline OR method OR approach OR challenges OR framework OR tool OR architecture)

We executed the search string in ACM, Scopus, and IEEE databases. This initial step followed a selection process using predefined inclusion and exclusion criteria. The inclusion criteria were defined to ensure the relevance of related studies to the scope of the research, while the exclusion criteria aimed to remove studies that did not meet the established objectives. The selection criteria are illustrated in **Table 2**.

Table 2. Inclusion and Exclusion Criteria for Study Selection.

Inclusion Criteria (IC):
IC01 - Studies related to *Interoperability* testing in Internet of Things applications. This criterion includes studies addressing concepts or practices of *Interoperability* testing applied to IoT systems.
IC02 - Studies presenting an IoT testing guide or similar artifacts. This criterion includes studies that focus on developing specific IoT testing guides or similar artifacts.
IC03 - Papers available in English or Portuguese. This category includes papers written in English, due to its widespread dissemination of the language within the scientific community, and in Portuguese, which is the native language of the researchers.
Exclusion Criteria (EC):
EC01 - Duplicate studies have only one version selected.
EC02 - Studies that do not address *Interoperability* testing for IoT applications: studies that do not present concepts, methodologies, or approaches for *Interoperability* testing in IoT applications.
EC03 - The study is not accessible.
EC04 - The study is a summarized version of another. In this case, the full version of the study has been selected.

Executing the search string in the databases returned 393 studies from ACM, 288 from Scopus, and 134 from IEEE.

[1] `https://parsif.al/`

3.2 Analysis of the Selected Studies

The second activity was the "Analysis of the Selected Studies". This analysis occurred in two rounds. In the first round, we read the titles and abstracts of the 815 identified studies and selected 102 preliminary studies. In the second round, we conducted a full reading of these studies: 50 from ACM, 24 from Scopus, and 28 from IEEE. Based on this set, the guide was filled out and structured with 50 studies.

The data extraction focused on crucial aspects for building the guide, including definitions of *Interoperability*, correlations among characteristics, specific testing challenges related to interoperability in IoT, and configuration requirements for test environments. We also analyzed the decomposition of *Interoperability* into subcharacteristics, the properties used to evaluate these subcharacteristics, related measures, test cases, cost-benefit aspects, and tools employed in the studies.

3.3 Construction of the Guide

The construction of the guide consisted of developing the content according to the 12 topics that form its structure, as described below.

1. **Definition of the characteristic.** Based on the literature review, this topic presents the main definitions of the *Interoperability* characteristic, with the purpose of establishing a standard of knowledge regarding this characteristic.
2. **Correlation of characteristics.** This topic aims to present the correlation of *Interoperability* with other IoT characteristics, allowing the identification of their impact, which can be positive, negative, or both.
3. **Challenges of interoperability testing.** This topic aims to present the challenges faced in testing the *Interoperability* characteristic in IoT applications.
4. **Test environment configuration.** This topic presents the configuration of a test environment, covering hardware and software required for validating the *Interoperability* characteristic.
5. **Definition of subcharacteristics.** This topic presents the definitions of the subcharacteristics of *Interoperability* identified in the literature, describing how these subcharacteristics were broken down for a more precise evaluation of the interoperability characteristic. The goal is to ensure conceptual standardization and a comprehensive understanding of each subcharacteristic.
6. **Contextualization.** This topic introduces a subcharacteristic in the context of IoT, explaining specific challenges it poses in ensuring effective interoperability. This topic also outlines the properties related to the *Interoperability* subcharacteristics.
7. **Abstract test cases.** This topic presents abstract test cases that can be concretized and adapted in the context of the interoperability testing process in IoT .
8. **Measurement.** This topic provides measures identified in the literature for evaluating *Interoperability* in IoT applications.
9. **Impact of subcharacteristics.** This topic evaluates the impact that one subcharacteristic may have on another, whether positive or negative.
10. **Cost-benefit.** The cost-benefit calculation for conducting interoperability testing is presented using specific variables. The calculation involves the relationship between the impact of the characteristic and the total cost of test execution.
11. **Suggested tools.** This topic suggests tools identified in the literature that can assist the test execution and the measurements.
12. **Example of using the guide.** This topic provides an example that demonstrates the step-by-step use of the *Interoperability* Testing Guide based on an IoT scenario.

Additionally, topics were enriched with examples, such as abstract test cases, measurements with explanations, device examples, and a detailed explanation of the test environment configuration.

3.4 Observation of the Behavior of IoT Applications

Interoperability testing plays a crucial role in ensuring the effectiveness of collaboration between IoT devices and systems. Observing IoT applications in real-world settings provides valuable *insights* into their interactions with the environment, their responses to variable conditions, and their ability to handle unforeseen events.

In the fourth activity, "Observation of the IoT Applications' Behavior", the guide was used in an IoT web application called Rottas UFC[2]. This application aims to support public transportation users by providing information such as bus schedules, route details, and real-time location through mobile devices with GPS.

In this context, the observation of the Rottas-UFC web application is particularly notable. When analyzing its performance in real-world situations, specific challenges related to *Interoperability* were identified, such as device resource limitations, communication between devices, and different protocols. Moreover, the guide's usage facilitated the creation of new test cases, focusing, for example, on analyzing protocol compatibility and verifying message consistency during device communication. Therefore, observing the actual behaviors of IoT applications is essential to ensure that the tests accurately reflect the challenges faced when integrating diverse systems.

3.5 Evaluation of the Interoperability Guide

The fifth activity conducted was the "Evaluation of the Interoperability Guide". We performed this evaluation in three stages, all designed to ensure the quality of the proposed guide. For guide evaluations, we selected Rottas UFC, a real-world IoT application developed by a postgraduate student within the GREat[3] (Group of Computer Networks, Software Engineering, and Systems) research group, where one of the authors is affiliated.

[2]The developers have authorized using the Rottas-UFC web application. It is currently being registered under number 512024000782-2. Available on: `https://rottas-ufc.github.io/webapp/#/`

[3]https://www.great.ufc.br/

In the first stage, a domain expert evaluated the guide. It is worth noting that the expert conducting the evaluation is the same one who initially proposed the guide's structure and instantiated it to Performance characteristic in IoT [Carvalho *et al.*, 2022]. Based on this expert evaluation, several improvements were suggested and incorporated into the guide, such as the addition of new test cases to cover more complex scenarios, the inclusion of detailed examples of challenges encountered in interoperability, the introduction of illustrative figures correlating relevant characteristics and measures, as well as the creation of a correlation table between measures, test cases, and evaluated subcharacteristics.

In the second stage, we conducted three evaluations of the guide. The first evaluation was based on the Technology Acceptance Model [Davis, 1989], while the second was conducted through a controlled experiment. We performed these evaluations with undergraduate and graduate students from the Software Verification and Validation course offered during the 2023.2 semester at the Federal University of Ceará (UFC). Following the controlled experiment, we carried out the third expert evaluation. Thus, the researchers and academic experts evaluated the guide. These evaluations provided valuable *insights* into the guide's effectiveness, contributing to further adjustments and continuous improvements to the guide.

Simultaneously with the guide evaluation stages, all the necessary instrumentation material for conducting the controlled experiment was developed[4], ensuring the accuracy and reliability of the results obtained.

3.6 Refinement of the Guide

The last activity was the "Refinement of the Guide". The goal was to enhance its usefulness and effectiveness. This step was conducted based on the analyses and evaluations in previous activities (subsections 3.4 and 3.5), incorporating the assessment of a real-world IoT application as well as feedback from both experts and participants who evaluated the guide.

Based on the results of the three evaluations, several improvements were implemented in the guide, including:

- **Addition of an initial description.** A detailed explanation of how to use the guide was included, facilitating its application for different user profiles.
- **Enhancement of the practical example.** The test conducted on an IoT application was further detailed, outlining the steps taken and the results obtained to clarify the evaluation and make it more replicable.
- **Detailed metrics.** We provided concrete examples for each metric, including an additional explanation to improve the interpretation of the evaluated criteria.
- **Inclusion of device usage examples.** The specific devices used in the tests were detailed, providing a more practical context for applying the guide.
- **Refinement of test cases.** New test cases were specified to cover a broader range of interoperability scenarios, ensuring a more comprehensive evaluation.
- **Adjustments to the cost-benefit formula.** The cost-benefit formula was revised according to feedback, improving accuracy and better reflecting the actual testing effort.
- **Improvement of guide illustrations.** We restructured the figures depicting the correlations between IoT characteristics to enhance visual clarity and facilitate data interpretation.

The aforementioned improvements were incorporated into the guide, and a consolidated version was provided, thereby enhancing its applicability and effectiveness for evaluating interoperability in IoT environments.

4 Interoperability Testing Guide

The interoperability testing guide provides a comprehensive approach for evaluating interoperability in different IoT application scenarios.

In the guide, the interoperability characteristic is broken down into four subcharacteristics: *Data Semantics*, *Communication Protocols*, *System Integration*, and *Network Protocols*. For each subcharacteristic, the guide provides topics related to their definitions, properties, test cases, and metrics.

The guide covers 12 topics, each organized as a section. The complete version is available in the repository of this study[5].

Next, we present the sections of the Interoperability Testing Guide.

4.1 Definition of Interoperability

This initial section of the guide aims to present the definitions of *Interoperability* characteristic. The goal is to standardize the interpretation of what will be tested, particularly for software engineers and professionals from different fields (e.g., Software Engineering, Computer Networks). The section titled "Definitions of Interoperability" presents six definitions of Interoperability extracted from the literature and standards such as ISO/IEC 21823 and ISO/IEC 30141 [ISO/IEC 21823, 2022; ISO/IEC 30141, 2018]. In the following, we present three examples of these definitions:

(1) Interoperability is *the ability of two or more systems or components to exchange information and effectively utilize the information exchanged* [Legner and Wende, 2006].

(2) Interoperability is *the ability for different systems, devices, or applications to communicate and interact efficiently and smoothly* [Ferreira, 2014].

(3) Interoperability in IoT *refers to the ability of connected devices (such as sensors, smart devices, gateways,* etc.*) to communicate and interact efficiently and transparently, regardless of their origin, manufacturer, or communication protocol* [ISO/IEC 21823, 2022].

4.2 Correlation of Characteristics

Recognizing and applying proper characteristic correlation is essential to identify potential conflicts that may arise in IoT

[4] `https://drive.google.com/drive/folders/1DozFXdxNVxTbI5fs3pI81TOuQ2JEclJU`

[5] `https://github.com/karinascb/INTEROPERABILITY-TESTING-GUIDE-FOR-IoT`

testing, as discussed in our previous work [Carvalho *et al.*, 2022]. In this study, a set of 27 IoT-related characteristics were identified through systematic mapping.

In the context of *Interoperability*, 14 IoT characteristics were identified and correlated [Branco *et al.*, 2024]. The section titled "Correlation of Characteristics" in the guide presents these correlation as illustrated in Figure 2.

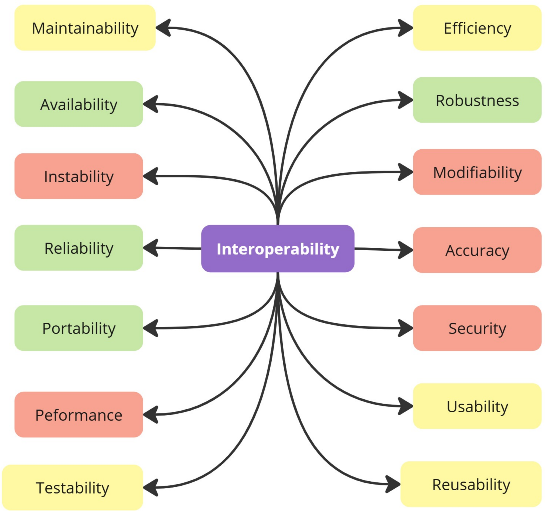

Figure 2. Correlations of IoT characteristics with Interoperability [Branco *et al.*, 2024].

Following the recommendation of Carvalho et al. [Carvalho *et al.*, 2022], the characteristics were grouped into three categories: *positive* (green rectangles), indicating a favorable influence on interoperability; *negative* (red rectangles), indicating the opposite; and *variable* (yellow rectangles), depending on the context of the application being tested.

As illustrated in Figure 2, the "Availability" characteristic can positively impact interoperability when there are enough servers to meet demand, accelerating the message exchange process. However, it can have a negative effect if this condition is not met.

The correlation analysis presented in Figure 2 was constructed to highlight which characteristics according to ISO standards-show influence on IoT interoperability. The figure includes both main characteristics, extracted from two main references: the ISO/IEC 25010 and the ISO/IEC 25012.

For example, ISO/IEC 25010 defines eight main quality characteristics: Acuracy, Performance, Efficiency, Usability, Reliability, Security, Avaliability, Maintainability and Portability. Some elements represented in the figure, such as *Testability* and *Modifiability*, are explicitly defined as subcharacteristics of *Maintainability*. These were included because they demonstrated specific influence on interoperability scenarios and were considered individually due to their direct influence on system integration.

In the case of *Testability*, its inclusion is justified by the importance of verifying whether interoperable components can be effectively tested regarding message exchange, supported protocols, and interface compliance. Systems with low testability tend to hinder the detection of integration faults, compromising interoperability.

Modifiability, on the other hand, was selected for representing the system's ability to quickly adapt to new standards, integration requirements, or changes in heterogeneous devices, a recurring demand in IoT ecosystems.

Thus, the selection of these characteristics was guided by theoretical foundations and their practical relevance observed in the literature, maintaining coherence with the guidelines of the ISO standards used as a basis.

Additionally, *Availability*, *Testability*, *Robustness* and *Instabillty* are not part of ISO/IEC 25010, but are defined in ISO/IEC 25012 characteristics of quality. Their presence in the figure reflects their strong impact on IoT environments, where real-time availability of services and precision in exchanged data are fundamental for successful device interaction.

Therefore, the 14 characteristics shown were not selected to replicate the ISO model, but rather to represent the practical influence each one has on interoperability in real-world IoT scenarios. This decision was supported by empirical findings and analysis.

Next, we present the definitions of these characteristics according to positive, negative, and variable correlations.

Positive correlation with Interoperability

1. **Availability.** Refers to the system's ability to be operational and accessible when needed, minimizing interruptions or failures. High availability favors interoperability by ensuring that different devices are always accessible for data exchange. On the other hand, low availability can compromise communication between systems, especially in distributed architectures.
2. **Robustness.** This focuses on the system's ability to deal with adverse conditions, such as unexpected inputs or failure situations, while maintaining its functionality. Robustness enhances interoperability by enabling the system to continue operating even in the event of failures in partner devices. In contrast, fragile systems may fail due to minor inconsistencies in communication protocols.
3. **Reliability.** This is related to the system's ability to perform its functions consistently and error-free over time. A reliable system increases confidence in the exchange of information between devices, which is essential for interoperability. However, recurring failures can compromise communication and hinder cooperation between heterogeneous systems.
4. **Portability.** Refers to the ability of a system to be easily transferred or adapted to different environments or platforms without significant loss of functionality. Portable systems tend to be more interoperable, as they can run on multiple platforms. Conversely, the lack of portability limits the use of the system in specific contexts, reducing its integration capacity."

Negative correlation with Interoperability

5. **Instability.** Refers to the tendency of the system to suffer unexpected failures or outages, resulting in inconsistent operation. An unstable system can interrupt the exchange of information between interoperable devices, causing synchronization failures or data loss, which makes it difficult for the expected joint operation.

6. **Modifiability.** This addresses specifically the ease of changing a system to meet new requirements or correct problems without causing undesirable impacts. According to ISO/IEC SQUARE 25000 [ISO/IEC 25000, 2005], modifiability is a subcharacteristic of maintainability.
7. **Accuracy.** It refers to the degree to which the system provides correct results in accordance with expected values. In the context of interoperability, accuracy is essential to ensure that data exchanged between different devices is interpreted correctly, avoiding inconsistencies in integration results. Failures in this aspect can compromise the reliability of cooperative operations, directly affecting the quality of communication between systems.
8. **Security.** This relates to the protection of system information and resources from unauthorized access, cyberattacks, and threats. Interoperability requires devices to exchange data continuously; in this context, well-implemented security measures ensure that this exchange occurs securely. However, overly restrictive security mechanisms can hinder or block legitimate integrations.
9. **Performance.** Refers to the system's ability to respond effectively to requests and operate within established limits, ensuring acceptable response times. Good performance contributes to interoperability by enabling fast responses between connected devices. On the other hand, unsatisfactory performance can cause delays in communication, compromising integration between components.

Variable correlation with Interoperability

10. **Maintainability.** This refers to the ability to change a system efficiently, including corrections or improvements, with low impact. Systems with high maintainability tend to adapt better to interoperability needs, as they can be modified to meet integration requirements with other systems. In contrast, systems with low maintainability face challenges in updating and adapting for compatibility. ISO/IEC 25000 [ISO/IEC 25000, 2005] defines testability and modifiability as subcharacteristics of maintainability.
11. **Usability.** This addresses the system's ease of use, ensuring it is intuitive and efficient for users. Good usability can facilitate the configuration and operation of interoperable devices for non-technical users. However, confusing or poorly designed interfaces can hinder the correct configuration of interoperability parameters.
12. **Reusability.** This refers to the ability of system components or modules to be reused in different parts of the software or other projects. Reusable components tend to follow standards, facilitating their integration with other systems. The absence of this characteristic can lead to the creation of isolated and poorly compatible solutions.
13. **Testability.** This relates to the ease of performing effective tests on the system to verify its quality and functionality. According to [Filho *et al.*, 2021], testability can assess software quality, as it reflects the ability to test a system with minimal effort. Indeed, high testability in systems allows for the effective validation of whether interfaces and communication protocols are working correctly, which is essential to ensuring interoperability. Conversely, difficulty in testing these interactions can hide integration failures.
14. **Efficiency.** This concerns the optimized use of resources such as CPU, memory and bandwidth to ensure adequate system performance. In the context of interoperability, efficiency plays a critical role in maintaining seamless communication between systems, especially in environments with limited computational or network resources. Inefficient use of resources can lead to delays or failures in data exchange, negatively impacting the integration and cooperation among heterogeneous devices.

4.3 Interoperability Testing Challenges

This topic presents the challenges that impact interoperability testing in IoT applications. A total of 20 challenges were identified. **Figure 3** shows these challenges, divided into three groups: (i) the most critical challenges mentioned in the literature (red lines), (ii) the most cited challenges in the literature (green lines), and (iii) challenges observed in practice in IoT applications (yellow lines).

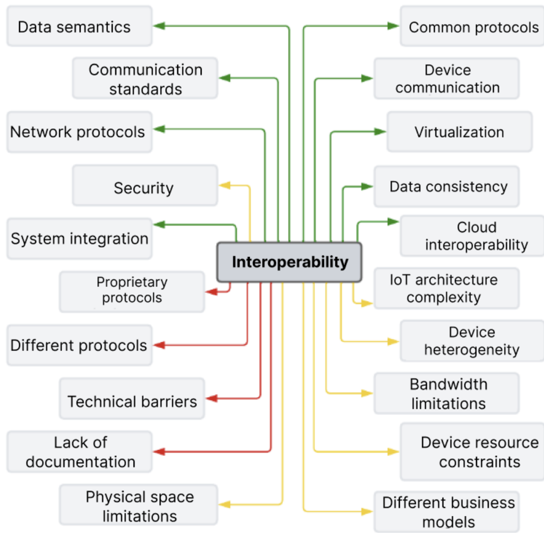

Figure 3. 20 challenges related to IoT interoperability testing [Branco *et al.*, 2024].

Below, we present four of the 20 interoperability challenges that impact the testing of IoT applications:

1. **Network protocols.** The challenge of testing compatibility between different network protocols (*e.g.,* TCP/IP, UDP, HTTP, and MQTT), involves addressing protocol diversity, integration complexity, and ensuring effective communication between devices from different manufacturers. The challenge is: *"How to test the compatibility between different network protocols?"* [Zanella *et al.*, 2014].
2. **Data semantics.** The challenge of ensuring the semantic consistency of data exchanged between IoT devices during testing, avoiding ambiguities, and ensuring a common interpretation of the data between devices and systems. The challenge is: *"How to test the semantic consistency of data exchanged between IoT devices?"* [Perera *et al.*, 2014].

3. **Security.** The challenges related to security involve testing the effectiveness of protection measures implemented, ensuring the integrity and confidentiality of sensitive data, and ensuring that interoperability does not introduce vulnerabilities. The challenge is: *"How to test interoperability without compromising the security of IoT devices?"* [Sicari *et al.*, 2015].
4. **Lack of Documentation.** The lack of clear and detailed documentation about devices and protocols can hinder integration and communication between devices, affecting interoperability. In this context, the testing activity must handle incomplete information, integration difficulties, and the need to develop test documentation. The challenge is: "How to deal with the lack of documentation on devices and protocols?" [Vermesan and Friess, 2011].

4.4 Test Environment Configuration

An effective testing environment for interoperability in IoT should include several basic components, such as one or more IoT devices, network infrastructure, actuators, and an application responsible for decision-making and command sending.

To complement the understanding of the IoT test environment configuration, **Figure 4** gives an overview of the IoT ecosystem, highlighting the relationships among devices, the network, and the application. The image illustrates the flow of interactions among sensors (responsible for collecting data from the environment), the network infrastructure (which enables communication between devices), the actuators (which perform physical actions in response to commands), and the application, which handles the decision-making process.

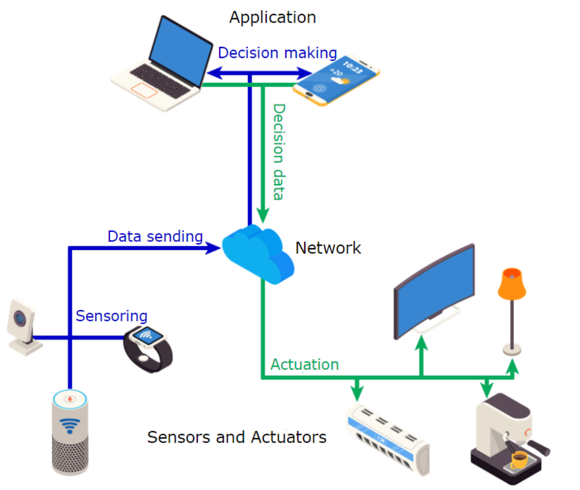

Figure 4. Relationship between sensors, actuators, network and IoT application [Carvalho, 2022].

Building a diverse environment that covers different devices, protocols, manufacturers, and network conditions is crucial to ensure that IoT devices can communicate effectively in real-world scenarios, regardless of their technical differences. The testing environment may include, for example, the following components to simulate interoperability:

IoT Devices. Evaluating interoperability in IoT requires using devices that can communicate with each other. These devices include sensors, control devices, and network devices compatible with the chosen communication protocol. For example, temperature sensors from manufacturers A, B, and C, as well as lighting actuators from manufacturers X, Y, and Z, may be used to ensure effective communication between different types of equipment.

Network Infrastructure. The network infrastructure must be set up to enable continuous and secure communication between IoT devices. This setup may involve wireless networks such as Wi-Fi and Bluetooth, switches, routers, and firewalls. Tests can be conducted on Wi-Fi, cellular, and LPWAN (Low Power Wide Area Network) networks to assess interoperability in various network environments.

Actuators. Actuators are essential devices that perform actions based on information received from sensors or commands from external systems. In the interoperability evaluation, actuators responsible for tasks such as opening and closing curtains, adjusting thermostat temperatures, and turning electronic devices on or off are important for testing system interaction and control.

Decision-Making and Command-Sending Applications. Applications or platforms that make decisions based on information provided by IoT devices and send commands to control them are crucial. These applications must be capable of communicating with devices from different manufacturers and protocols, ensuring system interoperability.

4.5 Subcharacteristics of Interoperabilty

In our approach, the *Interoperability* characteristic is subdivided into four subcharacteristics: *Data Semantics*, *Communication Protocols*, *System Integration* and *Network Protocols*. According to the instantiation methodology (see Section 3), if a selected characteristic has subcharacteristics, we should analyze and structure the data extracted from the literature review, beginning for the topics related to the subcharacteristics. In the guide, the following topics are presented for each subcharacteristic:

- **Definition of a Subcharacteristic.** This topic provides a formal description of a subcharacteristic based on ISO/IEC standards (*e.g.,* ISO/IEC 30141 [ISO/IEC 30141, 2018]) and clarifies its meaning within the context of *Interoperability*.
- **Contextualization.** This topic introduces a subcharacteristic in the context of IoT systems, explaining its importance and the specific challenges it poses in ensuring effective *Interoperability*. We also present the **properties** related to the *Interoperability*. A property is related to a quality measure element and can be quantified by a measurement method [ISO/IEC 25019, 2023]. In our approach, we defined measures based on the subcharacteristics' properties.
- **Abstract Test Cases.** This topic gives examples of abstract test cases that can be used to create concrete test cases. The test cases are hypothetical scenarios developed to assess the behavior of a subcharacteristic under various conditions.
- **Measurements.** This topic provides measures and evaluation methods to assess the effectiveness of the subcharacteristic's *Interoperability*, ensuring IoT devices

can operate correctly in different scenarios.

In the guide, we specified 25 abstract test cases for the four *Interoperability* subcharacteristics, as follows:

- 6 for *Data Semantics*;
- 7 for *Communication Protocols*;
- 5 for *System Integration*; and
- 7 for *Network Protocols*.

Each test case follows the same structure: the **title** describes the test; the **test environment** specifies the required configurations; the **precondition** establishes the necessary prior state; the **step-by-step** provides detailed instructions for execution; and the **postcondition** indicates the resulting state after completing the test.

In addition to the test cases, specific measurements were selected for each subcharacteristic. A total of 19 measures documented in the literature were selected to evaluate the *Interoperability*, as detailed below:

- 4 for *Data Semantics*;
- 5 for *Communication Protocols*;
- 5 for *System Integration*; and
- 5 for *Network Protocols*.

The following subsections present the four subcharacteristics of *Interoperability*, organized according to the above-mentioned topics.

4.5.1 Data Semantics

- **Definition.** According to ISO/IEC 30141:2018 [ISO/IEC 30141, 2018], *Data Semantics* refers to the meaning and context of data. It is the ability to interpret and understand the meaning of data, enabling different systems and devices to share and use data consistently and efficiently.
- **Contextualization.** In the IoT context, *Data Semantics* refers to the ability to ensure that IoT devices and systems share information efficiently and accurately, even when using different data formats or terminologies. This attribute is crucial when diverse IoT devices must collaborate to complete a specific task. We identified **eight properties** related to the *Data Semantics*, which are: *Common Interpretation*, *Terminology Harmonization*; *Data Mapping*; *Semantic Compatibility*; *Consistency*; *Accuracy*; *Vocabulary Comprehension*; and *semantic Coherence*.
- **Abstract Test Cases.** In the guide, we specified four abstract test cases for the

5 Wiki

The proposed guide was carefully structured into sections to simplify the interoperability testing process in IoT applications. To enhance usability and facilitate the application of the guide, a Wiki platform[6] was developed, inspired by the model proposed in a previous work [Carvalho *et al.*, 2022]. **Figure 5** presents the structure of the Wiki. The Wiki is designed as an interactive and user-friendly tool, focusing exclusively on the interoperability characteristic and providing a comprehensive environment to assist users in testing.

The Wiki allows users to intuitively navigate and select the correlated characteristics and subcharacteristics of *Interoperability*. Thus, users can select the subcharacteristics that best suit their applications' scenarios and specific needs. For each selected subcharacteristic, the Wiki dynamically displays the relevant properties, recommended metrics, and associated abstract test cases. This customization ensures users can efficiently configure their tests, aligning them with specific technical requirements or preferences.

Another relevant feature on the Wiki is the cost-benefit (CB). In the Wiki, users can calculate the CB related to the invested resources, such as time and effort allocated to conducting interoperability testing. This feature gives users a clear view of the trade-offs involved, aiding in strategic decision-making and more efficient resource allocation.

Additionally, the Wiki enables the generation of a customized test plan based on users' selections. This plan consolidates all relevant information into a structured and practical format, such as the chosen subcharacteristics, metrics, and test cases. The finalized plan can be exported and saved as a PDF file, ensuring users can access an organized and portable document for test documentation.

Below, we describe the Wiki's functionalities based on the guide's topics.

- **Characteristic definition.** The first step involves selecting the definitions of **Interoperability** characteristic in the Wiki. Users can consult predefined definitions to include in their test plans.
- **Characteristic Correlation.** This step involves selecting the correlated characteristics that impact on **Interoperability**. The Wiki provides the definitions of all 14 correlated characteristics (see Section 4.2).
- **Subcharacteristic selection.** Users can select which subcharacteristics related to Interoperability will be evaluated. The Wiki presents four subcharacteristics (see Section 4). Detailed definitions of each subcharacteristic are provided to facilitate decision-making.
- **Property selection.** Users can also select specific properties to be evaluated. If any property impacts others, the Wiki raises warnings to highlight this relationship and encourage the inclusion of correlated properties.
- **Test case recommendation.** Based on the selected properties, the Wiki suggests related abstract test cases, which are visually highlighted. Users can follow the recommendations or customize the list by adding or removing test cases.
- **Metric suggestion.** Besides test cases, the Wiki recommends appropriate metrics to evaluate the

[6] Available at: `https://laviniamatosof.github.io/interoperability_WIKI/`

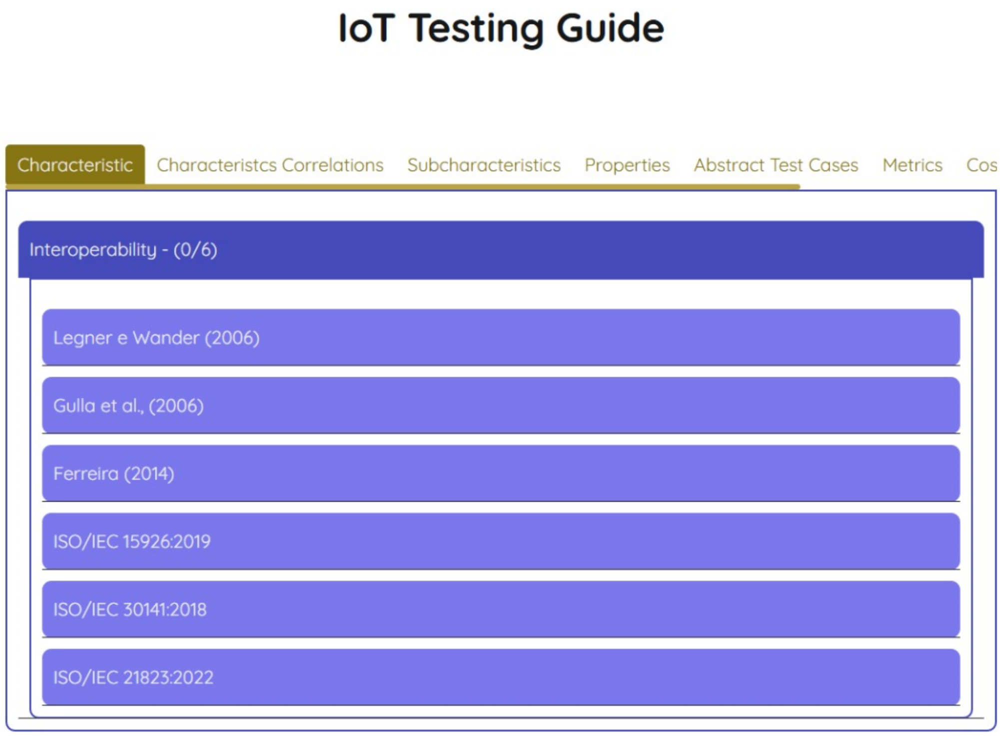

Figure 5. Overview of the Wiki designed for the Interoperability Testing Guide.

selected subcharacteristics based on the subcharacteristics' properties previously selected. Similar to test cases, the metrics can be adjusted according to the user's preferences.

– **Cost-benefit calculation.** The Wiki allows users to calculate the cost-benefit of their choices using average effort values. These values can be edited to reflect specific scenarios. The result is displayed in a quadrant that facilitates the analysis of the relationship between invested resources and expected benefits.
– **Tool suggestion.** The Wiki suggests appropriate tools for conducting the tests based on the selected properties, metrics, and test cases.
– **Test plan generation.** The Wiki allows users to generate a test plan based on their selections. Once completed, the interoperability test plan can be exported in PDF format, consolidating all information in an organized file.

6 TAM Model Evaluation

The Technology Acceptance Model (TAM) was developed by Davis [Davis, 1989] and is widely used to understand and predict user acceptance of new technologies. The TAM suggests that technology acceptance is primarily influenced by Perceived Usefulness (PU) and Perceived Ease of Use (PEOU).
However, to provide a more comprehensive analysis, the TAM model has been extended in several studies to include other relevant dimensions, as proposed by Venkatesh and Davis [Venkatesh and Davis, 2000; Venkatesh *et al.*, 2003]. In this research, in addition to the categories Perceived Usefulness (PU) and Perceived Ease of Use (PEOU), other additional categories were integrated: Future Use Intention (IU), Impact on Testing Efficiency (IET), and Overall Satisfaction (OS). The design of TAM evaluation is described below.

1. **Objective.** The evaluation conducted with the TAM Model aimed to analyze the acceptance and effectiveness of the *Interoperability* Testing Guide for IoT Applications, focusing on users' perception of the Guide's usefulness and ease of use.
2. **Context.** The guide evaluation was performed in a university-level course on Software Verification and Validation (V&V) offered to both undergraduate and graduate students. Before the evaluation, two leveling classes were conducted, each for two hours. The first class was theoretical, explaining the concept of *Interoperability* testing, and the second class was practical, performing *Interoperability* tests with smart devices, such as the Alexa virtual assistant. The proper evaluation by the students was then conducted during another face-to-face V&V class, involving six undergraduate students from a Computer Science course.
3. **Instrumentation.** The evaluation of the TAM Model dimensions was performed through a structured questionnaire. The categories of the model are:
   (a) **Perceived Usefulness (PU)** refers to the degree to which a person believes using a specific technology will improve their performance at work or in their activities. We formulated three questions to assess participants' perception of the usefulness of the *Interoper-*

ability Testing Guide.

(b) **Perceived Ease of Use (PEOU)** refers to the degree to which a person believes that using the technology will be effortless. Five questions were formulated to measure participants' perception of the Guide's ease of use.

(c) **Future Use Intention (IU)** assesses the likelihood that users will continue to use the technology in the future. We formulated two questions to assess whether participants intend to use the *Interoperability* Testing Guide in future projects.

(d) **Impact on Test Efficiency (IET)** measures how the use of technology influences the efficiency of tests performed. One question was formulated to assess the impact of the Guide on the efficiency of tests performed.

(e) **Overall Satisfaction (OS)** assesses users' satisfaction level with the experience of using the technology. To assess the level of general satisfaction of users regarding the *Interoperability* Testing Guide, we formulated three questions.

4. **Participants.** The evaluation using TAM involved six Computer Science students enrolled in the Software Verification and Validation course at the Federal University of Ceará (UFC). These students had been previously informed about the experiment but did not attend the initial leveling class for unspecified reasons. Therefore, they only participated in the TAM evaluation, without prior training. They were organized into pairs to evaluate the *Interoperability* Testing Guide for an IoT application. Five of the six participants were male, and their ages ranged from 22 to 26 years old.
Although these participants did not attend the initial leveling class, efforts were made to ensure the quality and consistency of their evaluation. Prior to using the guide, they received a brief contextualization about the objectives of the experiment, the scope of the IoT application under analysis, and the structure of the Interoperability Testing Guide. Despite the absence of formal training, the participants demonstrated basic familiarity with the concepts of software testing, as they were regularly enrolled in a course covering topics related. The decision to include them exclusively in the TAM evaluation phase was based on the intention to obtain an independent perspective on the guide's perceived usefulness and ease of use, aligned with the objectives of the TAM model. This approach contributed to capturing diverse user experiences, complementing the insights gathered from the initial experiment group.

5. **Results analysis.** The Likert [1932] scale was adopted as the response standard to analyze the results of the TAM questionnaire[7]. Its purpose is to specify the level of agreement or disagreement of respondents with respect to certain statements, where "strongly disagree" corresponds to 1, "neutral" to 3, and "strongly agree" to 5. The mode was used as a statistical measure to identify the most frequent value in the collected data.

[7]https://forms.gle/8j1zLyd4V4F9Uoi27

6.1 Results and Discussion

The analysis and results of the TAM model evaluation are presented below. The five stages assessed include Perceived Usefulness (PU), Perceived Ease of Use (PEOU), Future Use Intention (IU), Impact on Test Efficiency (IET) and Overall Satisfaction (OS). The evaluation was conducted with six participants, who were organized into pairs to evaluate the Rottas-UFC web application[8], using the *Interoperability Testing Guide*.

6.1.1 Perceived Usefulness

Three questions were developed to assess Perceived Usefulness (PU), as presented in **Table 3**.

Table 3. Questions to assess Perceived Usefulness.

ID	Question Description
PU1	Was the *Interoperability* Testing Guide useful for understanding how to conduct the *Interoperability* testing process in an IoT app?
PU2	Did the *Interoperability* Testing Guide prove effective in planning/specifying test cases in an IoT app?
PU3	Did the *Interoperability* Testing Guide facilitate the performance of tests in an IoT app?

Figure 6 shows the degree of agreement among participants regarding Perceived Usefulness of the *Interoperability* Testing Guide for performing tests in Rottas-UFC application. The analysis of these results provides a comprehensive view of the participants' perception regarding the effectiveness of the Guide in the various phases of the testing process.

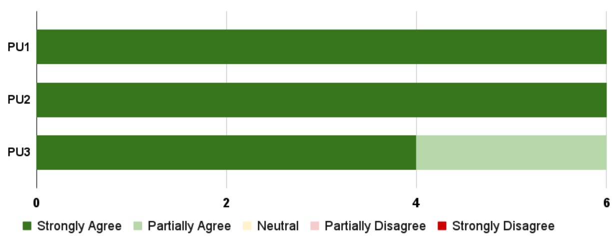

Figure 6. Usefulness Perception Assessment of the Guide.

When analyzing the results of question **PU1**, which assesses the usefulness of the Guide in understanding the process of conducting tests, we observed that all six participants (100%) expressed total agreement with the

[8]The Rottas-UFC web application was provided by the authors and is in the process of being registered under number: 512024000782-2

Guide's effectiveness in this regard. This result indicates a unanimous perception that the *Interoperability* Testing Guide was useful for understanding the process of executing tests in an IoT application.

The high agreement among participants remains in question **PU2**, which investigates the Guide's effectiveness in planning and specifying tests. All six participants (100%) fully agreed that the Guide was effective at this stage, reflecting a clear perception of the Guide's usefulness in the test planning process.

Regarding question **PU3**, which examines the Guide's effectiveness in executing tests, four participants (66.7%) fully agreed with the Guide's effectiveness, while the other two participants (33.3%) showed partial agreement.

The PU category mode was 5. The results presented in **Figure 6** show a high perception of the usefulness of the *Interoperability* Testing Guide among the participants, highlighting its effectiveness both in planning and conducting *Interoperability* tests in IoT applications. However, the results showed that when using the Guide to execute tests, two participants perceived partial effectiveness.

6.1.2 Perceived Ease of Use

Five questions were developed to assess Perceived Ease of Use (PEOU), as presented in **Table 4**.

Table 4. Questions to assess Perceived Ease of Use.

ID	Question Description
PEOU1	Are the document structure (topics) and instructions in the *Interoperability* Testing Guide easy to understand?
PEOU2	Was the learning curve for using the *Interoperability* Testing Guide smooth?
PEOU3	Was the organization of the topics and the sequencing of the contents clear?
PEOU4	I consider that using the Guide facilitates the planning/specification of *Interoperability* tests in IoT app.
PEOU5	I consider that using the guide facilitates the execution of *Interoperability* tests in IoT app.

Figure 7 shows the participants' agreement regarding the ease of use of the *Interoperability* Testing Guide when conducting tests in the Rottas-UFC application. This assessment is crucial to understand whether the Guide is intuitive and efficient, facilitating the planning and execution of *Interoperability* tests in IoT applications.

The first question (**PEOU1**) assessed whether the structure of the topics and the instructions in the Guide were easy to understand. The results show that four participants (66.7%) fully agreed with this statement, while two participants (33.3%) partially agreed.

The second question (**PEOU2**) investigated whether the learning curve for using the Guide was smooth. The results showed that three participants (50%) fully agreed, while the other three (50%) partially agreed with this statement.

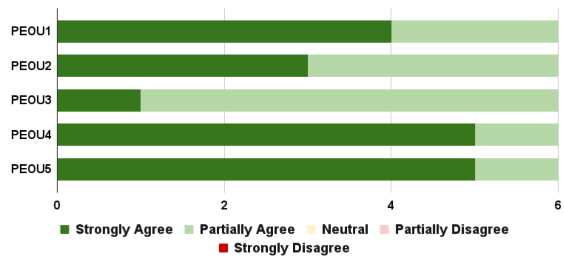

Figure 7. Perceived Ease of Use Assessment of the Guide

The third question (**PEOU3**) examined whether the organization of the topics and the sequencing of the contents were clear. The results show that one participant (16.7%) fully agreed, while another five participants (83.3%) partially agreed with this statement.

Regarding the perceived usefulness of the Guide, question (**PEOU4**) revealed that five participants (83.3%) fully agreed that using the Guide facilitates the planning and specification of *Interoperability* tests in IoT applications. Only one participant (16.7%) partially agreed.

The last question (**PEOU5**) addressed whether using the Guide facilitates the execution of *Interoperability* tests in IoT applications. In this question, five participants (83.3%) fully agreed with this statement, while one participant (16.7%) partially agreed.

The mode for the PEOU category was 5. However, the results indicated biased perceptions regarding the learning curve and the organization of the Guide's topics. Although the Interoperability Testing Guide was well received in terms of ease of use, the results suggest that improvements in the structure and clarity of the content could be beneficial. For example, participants reported that the organization of the topics could be more intuitive, suggesting that the content be restructured to facilitate navigation between sections. In addition, the difficulty in following the amount of textual and visual information presented was highlighted, which may have overloaded the reading experience. Some participants also felt that there was a lack of more varied scenarios in the test cases, which made it difficult to understand the applicability of the Guide in different contexts. Another recurring suggestion was to reduce the repetition of content, emphasizing focusing on practical examples that directly illustrate how to apply the tests. It was also recommended that direct links to the Guide be included in PDF format to facilitate quick access to instructions and step-by-step instructions during the execution of the tests.

The selection criterion adopted was participation in the leveling classes, which provided essential theoretical and practical knowledge about IoT and software testing. This choice aimed to ensure that all participants had a standard knowledge base before the assessment, allowing a more accurate analysis of the applicability of the Guide.

The assessment was carried out with two groups, totaling 12 participants. Although the sample size is small, this number is justified by the complexity and specificity of the test, which requires prior knowledge of IoT interoperability. In addition, the detailed and comprehensive content of the Guide may have contributed to the difficulties reported, especially regarding cognitive load and the need for greater objectivity in the instructions. Despite the limited sample size, the results provide valuable insights into the usability of the Guide and point out specific areas that can be improved to make it even more effective in supporting the planning, specification, and execution of interoperability tests in IoT applications.

6.1.3 Future Use Intention

For the analysis of Future Use Intention, we formulated two questions. They addressed the participants' willingness to use the Guide in future projects and to recommend it to colleagues. The questions are presented in **Table 5**.

Table 5. Questions to assess Future Use Intention.

ID	Question Description
UI1	Do I intend to continue using the *Interoperability* Testing Guide in future IoT test projects?
UI2	Would you recommend the *Interoperability* Testing Guide to colleagues who conduct tests on IoT app?

Figure 8 illustrates the participants' agreement regarding the Future Use Intention of the *Interoperability* Testing Guide. This assessment is critical to understanding the applicability and relevance of the Guide to your context or for use in academia and industry.

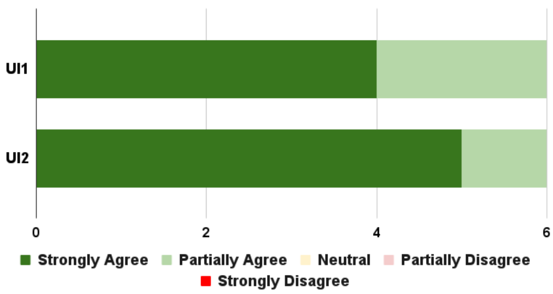

Figure 8. Future Use Intention of the Guide.

Question **UI1** revealed that four participants (66.7%) intend to continue using the *Interoperability* Testing Guide in future IoT testing projects. The other two participants (33.3%) partially agreed.
Question **UI2** revealed that five participants (83.3%) would recommend the *Interoperability* Testing Guide to colleagues who conduct tests on IoT applications. A single participant (16.7%) partially agreed.
The mode for the UI category was bimodal, with modes 4 and 5. The results demonstrate a high intent for continued use and recommendation of the *Interoperability* Testing Guide.

6.1.4 Impact on Test Efficiency

To analyze the impact on test efficiency, a question illustrated in **Table 6** was formulated to evaluate the contribution of the *Interoperability* Testing Guide to the overall effectiveness of tests in an application IoT.

Table 6. Questions to assess Impact on Test Efficiency.

ID	Question Description
IET1	Has using the *Interoperability* Testing Guide contributed to the overall effectiveness of testing an IoT app?

Figure 9 shows agreement among participants regarding the overall effectiveness of testing using the *Interoperability* Testing Guide.

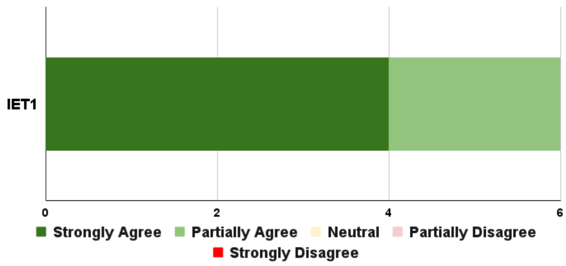

Figure 9. Impact on Test Efficiency of the Guide.

The results indicate that four participants (66.7%) strongly agreed that the Guide significantly contributed to the overall effectiveness of testing in IoT applications (**IET1**), while two participants (33.3%) partially agreed. The mode for the ITE category was 5. The results indicate a high perception that the *Interoperability* Testing Guide contributes significantly to the overall effectiveness of testing in IoT applications.

6.1.5 Overall Satisfaction

Three questions were developed to assess overall satisfaction. The questions are presented in **Table 7**.

Table 7. Questions to assess Overall Satisfaction.

ID	Question Description
OS1	Was the *Interoperability* Guide clear and understandable in terms of instructions and approach to test planning?
OS2	Was the use of the *Interoperability* Testing Guide clear and understandable in terms of instructions and approach to test execution?
OS3	Overall, am I satisfied with the experience of using the *Interoperability* Testing Guide?

Figure 10 shows the participants' agreement regarding the overall effectiveness of the tests performed using the *Interoperability* Testing Guide.
Regarding the clarity and comprehensibility of the *Interoperability* Testing Guide for test planning (**OS1**), four participants (66.7%) strongly agreed that the Guide

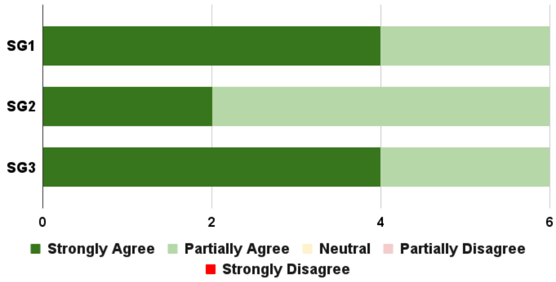

Figure 10. Overall Satisfaction with the Guide.

was clear and comprehensible, while two participants (33.3%) partially agreed with this statement.

For test execution (**OS2**), one participant (16.7%) strongly agreed, while five participants (83.3%) partially agreed with the clarity and comprehensibility of the Guide.

Regarding overall satisfaction (**OS3**), it was revealed that four participants (66.7%) strongly agreed that they were satisfied with the experience of using the Guide, while two participants (33.3%) partially agreed.

The mode for the OS category was 5. However, the results revealed a variable perception regarding the clarity and comprehensibility of the Guide. Although most participants considered the Interoperability Testing Guide clear for planning tests, the perceived clarity for executing tests was less uniform. The responses suggest that, while the Guide is well-rated for planning, improvements are needed in its clarity and comprehensibility during test execution to optimize the overall user experience.

7 Controlled experiment

In this study, we designed a controlled experiment [Wohlin *et al.*, 2012] to evaluate the effectiveness of the *Interoperability Testing Guide* in practice by analyzing its content and structure. This experiment was conducted in parallel with the TAM evaluation (see Section 6) and comprised nine activities, as shown below.

1. The **Experiment's Idea** focuses on evaluating the practical effectiveness of the Interoperability Testing Guide. The central aim was to compare the Interoperability tests conducted in the IoT application Rottas UFC between two participant groups: one using the guide and the other not. Participants conducted tests covering all stages from planning to execution and fault identification.
2. Regarding **Definition of the objective**, we assess whether the *Interoperability Testing Guide for IoT Applications* meets the proposed objective by evaluating practical application under controlled conditions to ensure an accurate analysis of the results.
3. In **Participant selection**, 12 students were invited to attend the first theoretical and practical class. These students were grouped into two groups: G1, composed of six participants who used the guide, and G2, composed of six participants who did not. This division enabled a precise comparison of participants who had used the guide and those who followed traditional procedures. The composition of the groups is detailed in **Table 8**.
   To ensure comparability between the groups, a preliminary leveling process was conducted with all participants. This process included presenting fundamental concepts related to the Interoperability characteristic and interoperability testing. Additionally, theoretical materials were provided, and supervised practical exercises were conducted in the classroom to promote a common understanding among participants prior to the experiment. Efforts were also made to maintain a balanced distribution between the groups in terms of educational level (undergraduate and master's), professional background (academic and/or industry), and prior experience with the relevant topics. In both groups, at least one participant had prior experience with interoperability testing, while the others demonstrated familiarity with non-functional requirements. This systematic approach helped to minimize biases related to technical knowledge and contributed to a more robust comparison between the groups.
4. The **Definition of research questions** (RQs) aims to guide the evaluation. The following questions were formulated accordingly:
   (a) **Q1. Does using the guide reduce the effort required to conduct *Interoperability* testing?** *This question investigates whether the guide facilitates the execution of interoperability tests, making the process more efficient and less laborious.*
   (b) **Q2. Does the guide facilitate the specification of a greater number of test cases for the evaluation of *Interoperability*?** *This question investigates whether the guide helps to increase the number of test cases specified, providing more extensive and detailed coverage in the evaluation of Interoperability.*
   (c) **Q3. Does the guide facilitate the identification of failures that are specific to certain protocols or configurations in an IoT application?** *This question examines whether the guide contributes to detecting specific failures related to certain protocols or configurations, which other testing techniques might not easily identify.*
5. Based on the research questions, the **Definition of hypotheses** were established to verify Q1, Q2 and Q3 questions. The hypotheses were divided into null ($\mathbf{H_0}$), which states that there is no significant relationship, and alternative ($\mathbf{H_1}$), which indicates that a significant relationship exists, as illustrated in **Table 9**.
   In this research, the null hypotheses aim to verify whether the approach based on an Interoperability Testing Guide requires the same testing effort and detects the same number of failures as traditional tests. Conversely, the alternative hypotheses clarify that structured guides reduce testing effort,

Table 8. Profile of the Experiment Participants.

Education	Position	Expertise
Group G1 - With the Guide		
Master	Academic and Industry	*Interoperability*
Master	Academic and Industry	Non Functional Requirement
Undergraduate	Academia	*Interoperability*
Undergraduate	Academia	Non Functional Requirement
Undergraduate	Academia	Non Functional Requirement
Undergraduate	Academia	Non Functional Requirement
Group G2 - Without the Guide		
Undergraduate	Academia	*Interoperability*
Undergraduate	Academia	Non Functional Requirement
Undergraduate	Academia	Non Functional Requirement
Undergraduate	Academia	Non Functional Requirement
Undergraduate	Academia	Non Functional Requirement
Undergraduate	Academia	Non Functional Requirement
Undergraduate	Academia	*Interoperability*

Table 9. Relationship between Hypotheses and RQs.

Research Question	Hypothesis
Q1	$\mathbf{H_{0,0}}$, $\mathbf{H_{1,1}}$
Q2	$\mathbf{H_{0,1}}$, $\mathbf{H_{1,2}}$
Q3	$\mathbf{H_{0,3}}$, $\mathbf{H_{1,3}}$

increase the number of generated test cases, and the number of identified IoT failures compared to traditional techniques. The description of each hypothesis is presented in **Table 10**.

To evaluate the null hypotheses, we applied the *Student's t-test* proposed by Fienberg [Fienberg, 2006]. The decision to reject or accept the null hypotheses is based on the p-value *(probability value)* relative to the established significance level. The null hypothesis is rejected if the p-value is less than the defined significance level. A significance level of 0.05 is commonly used in statistical analyses.

The validity of applying the *Student's t-test* was previously verified by checking the required statistical assumptions, the results[9] are available in the repository of this research[10]. Normality within each group ($n = 3$) was assessed using the Shapiro–Wilk test [Shapiro and Wilk, 1965], which is particularly recommended for small samples ($n < 50$). The results indicated an approximately normal distribution ($p > 0.05$), especially for the variables "Effort: Time spent (in minutes) on planning" and "Number of test cases". Homogeneity of variances between the groups was evaluated using Levene's test [Levene, 1960], which also showed no significant differences ($p > 0.05$). Based on these results, the application of the *Student's t-test* was considered appropriate, even with a small sample size, as supported by [Ghasemi and Zahediasl, 2012] and [Kim, 2015], who emphasize the robustness of the *Student's t-test* under such conditions when assumptions are satisfied. To strengthen the reliability of the analysis, the non-parametric Mann–Whitney U test [Mann and Whitney, 1947] was also applied as a sensitivity check. The results were consistent with those of the *Student's t-test*, reinforcing the robustness of

Table 10. Experimental Hypotheses on *Interoperability* Testing.

Hypothesis	Description
Null Hypotheses	
$\mathbf{H_{0,0}}$	The structured guide-based approach to conducting *Interoperability* testing activities requires the same amount of effort as traditional testing.
$\mathbf{H_{0,1}}$	The structured guide-based approach produces the same number of test cases as traditional testing.
$\mathbf{H_{0,3}}$	The structured guide-based approach detects the same number of IoT faults as traditional testing.
Alternative Hypotheses	
$\mathbf{H_{1,1}}$	The structured guide-based approach reduces the testing effort more than traditional testing. ***Effort (With guide) < Effort (Without guide)***
$\mathbf{H_{1,2}}$	The structured guide-based approach produces more test cases than traditional testing. ***Number of test cases (With the guide) > Number of test cases (Without the guide)***
$\mathbf{H_{1,3}}$	The structured guide-based approach finds more IoT faults than traditional testing. ***Number of IoT failures (With the guide) > Number of IoT failures (Without the guide)***

[9] Statistics Kingdom was used as a support tool for carrying out statistical calculations, available at: https://www.statskingdom.com/index.html

[10] `https://github.com/karinascb/INTEROPERABILITY-TESTING-GUIDE-FOR-IoT`

the conclusions. However, for the variable "Number of faults detected", which exhibited a more asymmetric distribution, the Mann–Whitney test was prioritized in the interpretation of results. Considering the applicability of *Student's t-test* with significance level ($p - value > 0.05$), in this research, the testing effort is defined by the time spent on test planning. **Table 11** presents the hypotheses and the evaluation methods.

Table 11. Hypothesis Evaluation Methods.

Hypothesis	Evaluation Method
$H_{1,1}$	The time spent on planning will be counted, including defining test cases, selecting measures, and configuring the test environment. The average time for each group will be calculated. Data will be collected and compared between groups to assess whether there are differences in the effort required to conduct testing based on whether or not the guide is used.
$H_{1,2}$	The number of test cases generated by each group will be counted, and the average will be calculated for each. These results will be compared between groups to assess possible differences in testing effectiveness.
$H_{1,3}$	The number of IoT-related failures detected by each group will be counted, and the average will be calculated for each. These values will be compared to determine whether there are differences between groups in the number of failures identified using the guide.

6. In the **Application definition** activity, the IoT application, Rottas[2], was used. In this application, users can view routes and schedules for intracampus and intercampus buses, track the location of buses in real-time, check estimated arrival times at stops, obtain accurate information about the location of intracampus stops, search for specific locations within the UFC campus, such as buildings, colleges, laboratories and libraries, and filter and sort buses on a specific line based on the stop and/or schedule, facilitating efficient trip planning and reducing waiting time at bus stops. **Figure 11** presents screenshots of Rottas UFC, providing a visual overview of previously described functionalities.
The experiment was carried out *in loco* in the computer laboratory of GREat, with the participation of ten undergraduate and two graduate students under the supervision of the research advisor (second author).
7. A **Definition of instrumentation** included the preparation of fault report (a spreadsheet template), test execution script, video lessons related to *Interoperability* testing, and face-to-face leveling classes.
In the initial stage of the instrumentation, a repository with explanatory videos on the concept of *Interoperability* and on the execution of tests in IoT applications was made available. To level the students' knowledge, we organized two classes: a theoretical one, covering the fundamentals of *Interoperability* testing, including specific articles and tutorials in addition to presenting the main testing methodologies in IoT applications, and a practical one, in which smart devices were used, such as lamps and the virtual assistant Alexa.
The planning of the classes aimed to ensure uniform knowledge among all participants, allowing everyone to understand both the concept and the practice of *Interoperability* testing. In addition, all the templates developed were made available to all participants[10]. For the members of G1 (the group that used the guide), a PDF version of the *Interoperability* Test Guide for IoT Applications was provided.
To record the application failures identified by the participants, we provided a set of documents: (i) a template for reporting failures, (ii) definitions of the characteristics, and (iii) a definition of the severity levels of the failures. The template for reporting failures, based on the model proposed by Carvalho et al. [Carvalho *et al.*, 2022], includes the following fields:
(a) Description (description of the detected failure)
(b) Test case/measurement (test case or measurement in which the failure was detected)
(c) Characteristics (characteristics related to the detected failure)
(d) Severity (severity level of the detected failure).
This set of documents was developed to ensure systematic and detailed documentation of failures. To assist in filling out the failure report, we provided a template that includes the definition of the characteristics related to *Interoperability*, the definition of the specific IoT characteristics, and the description of the severity levels of the failures. In addition, the document contains a template for the failure report.
To collect data from the experiment, such as duration, documents used, and feedback, we administered a questionnaire to participants at the end of the experiment[10]. The final questionnaire for participants in group G1 (which used the guide) included more questions specifically aimed at its use. The distribution of these questions is presented in **Table 12**.
In addition to the artifacts mentioned above, the access link to the Rottas UFC application was made available to participants, requesting that each participant prepare a test plan and complete the *checklist* and the failure report. All instruments used in the experiment are available in the repository

Figure 11. Rottas UFC *web* application.

Table 12. Categories and Number of Questions Related to the Experiment.

Category	General questions	Additional questions	Sample additional questions
General	3	1	Was the process suggested in the guide for conducting the tests followed?
Artifacts	4	1	-
Execution	1	8	Based on the guide, was the test plan generated only once?
Effort	14	5	How much time was spent understanding the guide?
Difficulties	4	2	Did you have any difficulties with the tools suggested in the guide?
Final Considerations	2	-	-

github.
Although no pilot was conducted, we leveraged the instrumentation and insights from our previous evaluation of the Performance Testing Guide [Carvalho *et al.*, 2022]. We reviewed all artifacts and tested them with the Rottas UFC application to improve instrumentation.

8. For the **Execution of the experiment**, the participants started the process by individually answering a pre-test questionnaire. This questionnaire was intended to assess the participants' understanding of *Interoperability* testing and to ensure a "class leveling" in prior knowledge on the topic. Each step is detailed below:
   (a) **Step 1 - Understanding *Interoperability*.** For further clarification on *Interoperability* and IoT testing, participants received a presentation on *Interoperability* testing and watched related videos.
   (b) **Step 2 - Reading the Guide.** Participants in group G1 were instructed to read the guide fully to understand the concept of *Interoperability* testing for IoT applications.
   (c) **Step 3 - Understanding the Application.** Participants were instructed to learn about the Rottas UFC application, which would be the application under test.
   (d) **Step 4 - Application Configuration.** Participants were instructed to access the Rottas UFC application and perform their configuration according to the instructions provided.
   (e) **Step 5 - Access the Guide again.** Participants who used the guide were instructed to access it again for more information about *Interoperability*.
   (f) **Step 6 - Test Planning and Execution.** Participants were instructed to plan, build, and execute *Interoperability* tests as instructed.
   (g) **Step 7 - Filling out the Fault Report.** As faults were detected, participants were asked to complete the report (the spreadsheet) according to the template provided.
   (h) **Step 8 - Answering the questionnaire.** Upon completing the evaluation, participants were instructed to answer the questionnaire about the evaluation, filling out the three available tabs.
   (i) **Step 9 - Final Artifacts.** Participants were instructed to generate three artifacts at the end of the test: Test Plan, Failure Report, and Final Questionnaire.
   (j) **Step 10 - Sharing the generated artifacts.** Participants were instructed to send all documents generated during the evaluation to a specific email.

   Each participant had, on average, two hours to execute the experiment. The G1 participants who used the guide followed 10 steps, while the others followed 8. Steps 2 and 5 are exclusive to the G1 group that used the guide.
9. After executing the experiment, the **Analysis and Results** of the artifacts sent by the participants of each group was performed. The results were compared between the groups and discussed to

assess the contribution of the guide to improving *Interoperability* testing.

In the following subsection, details of the data obtained, the statistical calculations performed, and the results of the hypotheses will be presented, as well as the answers to the research questions listed for the experiment.

7.1 Analysis and results of the controlled experiment

Table 13 shows experiment results, separating groups that used the guide (G1) from those that did not (G2). For each participant, ID, planning time, test cases specified, and IoT failures reported are listed.
The time in minutes shows the effort spent reading the guide, setting up the test environment, preparing test cases, selecting measures, running tests, and defining the test plan scope.

Table 13. Results of the controlled experiment.

ID	**Time** (min)	**Test Case** (#)	**IoT failure** (#)
Group 1 - G1			
G1-1	50	8	3
G1-2	45	10	2
G1-3	40	6	0
Group 2 - G2			
G2-1	90	4	0
G2-2	60	3	0
G2-3	50	6	2

When analyzing **Table 13**, we observe that the group that used the guide (G1) produced more test cases and identified more IoT failures than the group that did not use the guide (G2). However, the planning and execution time were shorter for group G1, except for participant G1-1, whose planning time was equivalent to that of participant G2-3 in group G2. This difference can be attributed to the doubts expressed by some participants during planning and IoT application testing.
Although all the necessary support was previously provided, including face-to-face guidance and classroom discussions, these doubts may reflect a perception of complexity in conducting *Interoperability* tests. When faced with the need to start test planning from scratch, participants in the group without the guide may have felt unfamiliar with the procedure. Participants in the G1 group had access to the abstract test cases provided by the guide, which facilitated the process. In contrast, the G2 group had to create everything from scratch, further contributing to the observed difference.

7.2 Analysis and results of the research questions

This section presents the analyses and results of the research questions formulated to evaluate the effectiveness of the *Interoperability* testing guide. The research questions were designed to investigate specific aspects of the application of the guide and its impact on the conduct of the tests. For each question, hypotheses were formulated and tested through statistical analyses based on online tools[11][12][13] on the data collected during the experiment.
Q1 - Does using the guide reduce the effort required to conduct *Interoperability* tests? To answer this research question, two hypotheses presented in **Table 14** were evaluated.
We applied the *Student's t-test* to evaluate the null hypothesis with a significance level of 0.05. The data used consisted of the test-planning times, in minutes, collected from participants' responses to the final questionnaire of the experiment. The time values were then compared between groups G1 (with a guide) and G2 (without a guide).
The average time was calculated by adding the times spent on test planning for each group and dividing the total by the number of participants. For example, the values were added and divided by six. This analysis allowed us to verify whether there is a statistically significant difference between the test planning times of the two groups, which could indicate the effectiveness of the *Interoperability* test guide.
This result indicates a **statistically significant difference**, which means there is sufficient evidence to reject the null hypothesis $H_{0.0}$ - the structured guide-based approach to conducting *Interoperability* testing activities requires the same testing effort as traditional *Interoperability* testing.
The **statistically significant difference** implies that the difference observed in test planning times between groups G1 (with guide) and G2 (without guide) is not due to chance but rather to the effect of the intervention, which in this case is the use of the *Interoperability* Test Guide. In other words, the use of the structured guide demonstrated greater effectiveness in terms of reducing the effort required for planning *Interoperability* tests compared to traditional methods, as illustrated in **Table 14**.
As detailed when we presented the hypotheses, we compare whether the Effort (with the guide) is lower than the Effort (without the guide) (alternative hypothesis $H_{1,1}$), considering that the effort is measured by the time spent (in minutes) on testing planning.
Q2. Does the guide facilitate the specification of a larger number of test cases to evaluate *Interoperability*?
Table 15 shows the results for the null hypothesis to answer this research question.
We again applied the *Student's t*-test with a significance level of 0.05 to evaluate the null hypothesis. The data used were the number of test cases generated, calculated by dividing the groups. The p-value obtained was 0.023, representing a lower value than the significance level established at 0.05, resulting in a **statistically significant**

[11] `https://www.graphpad.com/quickcalcs/pvalue1.cfm`
[12] `https://pt.symbolab.com/calculator/statistics/p-value`
[13] `https://www.omnicalculator.com/pt/estatistica/teste-t`

Table 14. Results of hypothesis $H_{1,1}$.

Hypothesis	Comparison	G1 Mean	G2 Mean	Probability	Result
$H_{0.0}$ and $H_{1.1}$	Effort: Time spent (in minutes) on planning	45	66	p-value = 0.035	Statistically significant difference

difference. This result indicates that the null hypothesis $H_{0,1}$, which states that the structured guide-based approach produces the same number of test cases as traditional testing, should be rejected.

Q3 - Does using the guide improve IoT fault detection?

To answer this question, we evaluated the null $H_{0.3}$ and the alternative $H_{1.3}$ hypotheses (see **Table 10**).

We used the data on the number of detected IoT failures, as presented in **Table 13**. The p-value was 0.03 (see **Table 16**), which is below the significance level of 0.05, indicating a **statistically significant difference**. Therefore, the null hypothesis $H_{0,3}$, which states that the structured guide-based approach to *Interoperability* testing detects the same number of IoT failures as traditional *Interoperability* testing, should be rejected.

During the analysis of the failures identified by participants, duplicates were removed, ensuring each distinct failure was counted only once. In a second step, when the same failure was identified by participants across groups, it was counted once per group. For example, the group with the guide (G1) identified five failures, while the group without the guide (G2) identified two failures that G1 had already identified. As a result, we have five occurrences of failures in the experiment, but only three distinct failures in the system under test. For statistical analysis purposes, especially when evaluating the effectiveness of failure identification, we considered five **major** failures, since each group was evaluated individually.

8 Evaluation with experts

To ensure the precision, relevance, and practical applicability of the *Interoperability* Testing Guide for IoT Applications, we evaluated it with experts in Software Quality and Testing. The goal was to obtain technical and methodological validation of the guide, identify its strengths and potential improvements, and ensure it meets the needs and expectations of researchers and professionals in the Software Testing field. The methodology used in the evaluation conducted with the experts is presented below.

1. **Objective.** To evaluate the structure and content of the *Interoperability* Testing Guide from the perspective of experts in Software Quality and Testing.
2. **Context.** The evaluation was conducted remotely over two weeks using the Rottas UFC web application. A detailed script was provided to guide the process.
3. **Participants.** Experts with experience in both academia and industry. **Table 17** presents the experts' profile.
   We observe that out of the six experts, three work predominantly in the industry (ESP1, ESP5, and ESP6), while three work in both industry and academia (ESP2, ESP3, and ESP4).
   All experts selected for the evaluation demonstrated solid knowledge in software testing and non-functional requirements, particularly in the context of IoT systems. The selection aimed to ensure diversity in professional backgrounds and perspectives. The experts were invited to remotely evaluate the Interoperability Testing Guide, and all received prior instructions regarding the scope and objectives of the evaluation. Additionally, they were provided with a brief explanation of the context in which the guide was developed.
4. **Instrumentation.**The data collection instrument employed in the study was a questionnaire comprising 32 questions, organized into four distinct sections:
   - 6 questions about the expert's profile;
   - 7 questions related to the guide's structure;
   - 10 questions related to the guide's content;
   - 9 questions regarding the general perception of the use of the guide.

The responses were gathered using a Likert scale [Likert, 1932], ranging from 1 to 5, where 1 corresponds to " totally disagree" and 5 to "I totally agree." This scale was applied in Section "Evaluation of the Structure of the *Interoperability* Testing Guide "and Section 8.1.3 "General Perception of the Use of the Interoperability Testing Guide".

The experts assessed elements such as the clarity of the provided materials, the organization of the sections, and the guide's effectiveness in supporting test activity planning.

In the next subsections, we present the results of the experts' evaluation.

8.1 Results and Discussion

The results of the evaluation carried out with the experts are presented in three stages: (i) Evaluation of the structure and content of the *Interoperability* Testing Guide, as described in Section 8.1.1; (ii) Evaluation of the use of the *Interoperability* Testing Guide, detailed in Section 8.1.2; and (iii) General perception about the use of the *Interoperability* Testing Guide, presented in Section 8.1.3.

Table 15. Results of hypothesis $H_{1,2}$.

$H_{x,y}$	Comparison	G1 Mean	G2 Mean	Probability	Result
$H_{0,1}$ and $H_{1,2}$	Number of test cases/ Number of groups	8	4	p-value = 0.023	Statistically significant difference

Table 16. Results of hypothesis $H_{1,3}$.

$H_{x,y}$	Comparison	G1 Mean	G2 Mean	Probability	Result
$H_{0.3}$ and $H_{1.3}$	Number of faults detected/Number of groups	1.66	0.66	p-value: 0.03	Statistically significant difference

Table 17. Experts and their profiles.

ID	Education	Academic Title	Experience	Position	Expertise
ESP1	Computer Engineering	Bachelor	Professional	Test Analyst	Software, IoT and Interoperability testing
ESP2	Computer Science	Master	Academia and Industry	Test Analyst	IoT testing
ESP3	Computer Science	Bachelor	Academia and Industry	Test Leader	Software, IoT and Interoperability testing
ESP4	Data Science	Bachelor	Academia and Industry	Test Analyst	Software Testing
ESP5	Systems Analysis	Bachelor	Professional	Test Analyst	Software Testing
ESP6	Software Engineering	Master	Professional	Senior Test Analyst	Software, IoT and Interoperability testing

8.1.1 Evaluation of the Structure and Content of the Interoperability Testing Guide

To assess the structure of the guide, we formulated seven specific questions, as detailed in **Table 18**. These questions were designed to assess the suitability of the materials and the organization of the topics.
Figure 12 presents the agreement among experts regarding the evaluation of the structure and content of the guide. Each line represents a question related to the evaluation.

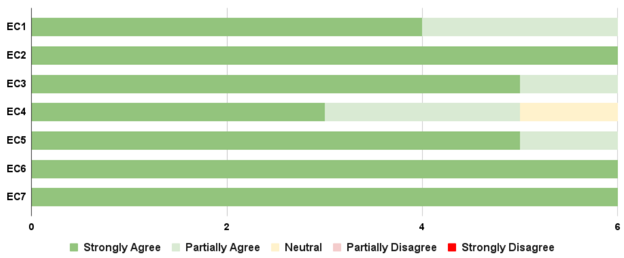

Figure 12. Evaluation of the structure and content of the guide.

For question **EC1**, about the adequacy and effectiveness of materials and instruments provided to understand the guide, the results indicated that four experts (66,67%) fully agreed with this statement. In contrast, two participant (33,33%) partially agreed.

Table 18. Assessment of the structure and content of the guide.

Code	Question
EC1	Were the materials and instrumentation provided for using the guide adequate and effective for understanding?
EC2	Did the organization of the topics of the *Interoperability* Testing Guide facilitate its understanding and use?
EC3	Were the topics covered in the guide organized logically and sequentially?
EC4	Were the abstract test case examples provided in the *Interoperability* Testing Guide useful for creating concrete test cases?
EC5	Did the *Interoperability* Testing Guide prove effective in planning *Interoperability* testing activities?
EC6	Were the examples and illustrations provided in the guide useful for understanding the concepts presented?
EC7	Did you consider that the guide covers all relevant aspects of *Interoperability* testing in an IoT application?

Regarding questions **EC2**, **EC6**, and **EC7**, all six experts (100%) fully agreed with the questions. In questions **EC3** and **EC5**, five experts (83.3%) expressed complete agreement. However, in question **EC4**, the distribution of responses was more varied: three experts (50%) fully agreed, two experts (33.3%) expressed partial agreement, and one expert (16.7%) expressed a neutral opinion.
We calculated the mode of the experts' responses. The mode for the *structure and content of the guide* was 5, indication strong agreement. The results suggest that the guide is well-structured and organized. However, question **EC4** revealed that despite overall positive feedback, some experts recommended improvements in certain topics, particularly regarding the usefulness of the examples provided for creating concrete test cases.

8.1.2 Evaluation of the use of the *Interoperability* Testing Guide

To evaluate the use of the *Interoperability* Testing Guide, we formulated 10 open-ended questions to investigate the effectiveness and applicability of the Guide in different stages of the testing process. The questions are presented in **Table 19**.

Table 19. Evaluation of the use of the *Interoperability* Testing Guide.

ID	Description
UG1	How was the execution of the tests?
UG2	How much time was spent reading and understanding the Guide?
UG3	How much time was spent on the entire evaluation, considering it consisted of five effort-intensive steps: (i) Reading and understanding of the Guide; (ii) Test planning; (iii) Environment configuration; (iv) Test specification; and (v) Test execution?
UG4	Did you use any tool suggested in the Guide? Which one(s)?
UG5	Did you use any tool that was not mentioned in the Guide? Which one?
UG6	Did you use all the topics in the Guide to plan, specify, and execute the tests?
UG7	Did you use any measurements? If so, how many were collected?
UG8	Did you use any measurements that were not presented in the Guide?
UG9	What IoT failures did you identify through the use of the Guide?
UG10	If you found failures in the tested application, did you have difficulty classifying and assigning severity to these failures using the template provided in the instrumentation?

The results obtained by the experts regarding the test execution (**UG1**) revealed that five experts (ESP1, ESP3, ESP4, ESP5, and ESP6) performed the tests exclusively manually, which represents (83,33%) of the participants. Only one expert (ESP2) used both manual and automatic approaches.
The prevalence of manual execution can be attributed to the complexity of the tests, which often require direct human intervention. Although automation offers efficiency, it can be more complex and time-consuming. The only exception was the expert ESP3, who combined manual and automatic methods. This choice highlights the advantage of integrating both methods for a more comprehensive and detailed evaluation, balancing the accuracy of manual testing with the efficiency of automation. It should be noted that ESP3 is a test leader with experience and knowledge in *Interoperability* testing.
Figure 13 presents the results obtained by experts on the *reading and understanding of the Guide* (**UG2**). This step is crucial to ensure that experts fully understand and follow the instructions provided in the Guide.

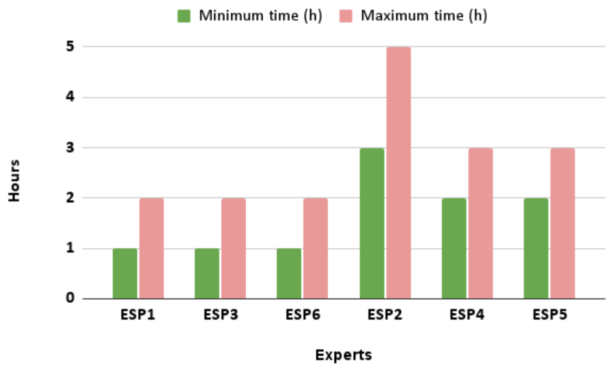

Figure 13. Time spent reading and understanding the Guide.

The results show a relationship between domain expertise and the time required to read and understand the Guide. Experts who work exclusively in software testing, such as ESP4 and ESP5, need between 2 and 3 hours to complete this step. In contrast, ESP1, ESP3, and ESP6, who have experience in Software, IoT, and Interoperability testing, completed the Guide in under 2 hours, suggesting that experience in these areas facilitates efficient comprehension of the Guide.
Conversely, the expert (ESP2), who has a background in academia and industry and a focus on IoT testing, took between 3 and 5 hours to complete the reading, indicating a more detailed and careful analysis or possible specific difficulties in understanding the material.
Variations in the time needed to read and comprehend the Guide may be associated with the experts' knowledge domains. Those with experience in IoT, software, and Interoperability testing generally took less time, while experts who focused exclusively on software testing took longer. However, the fact that most experts completed the reading and comprehension relatively quickly indicates that the Guide is generally well-structured and easy to understand, regardless of their domain knowledge. **Figure 14** presents the total time spent on the experts' evaluation (**UG3**), considering the five steps: reading and understanding the Guide, planning the tests, configuring the environment, specifying and executing the tests.

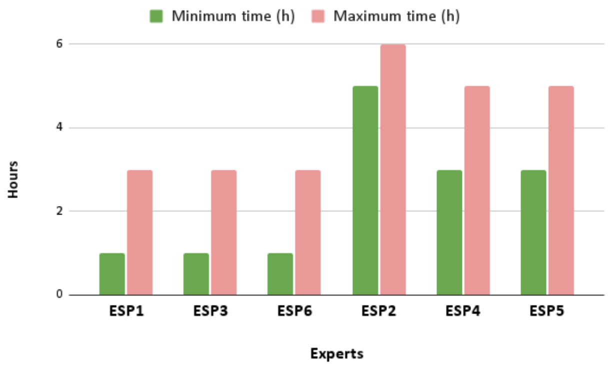

Figure 14. Total time spent by the experts during the evaluation.

The results show variation in total time spent on the evaluation, reflecting differences in the experts' knowledge domains. Experts working in software testing, IoT testing, and Interoperability testing, such as ESP1, ESP3, and ESP6, completed all steps in 3 hours. This result suggests that experience in these areas facilitates an efficient and well-planned approach to all steps of the assessment. In contrast, the ESP2 expert, who has experience only in IoT testing, spent 5 hours or more completing the assessment. This longer time may indicate a more detailed approach or possible specific execution difficulties.
Experts with experience only in software testing, such as ESP4 and ESP5, took between 3 and 5 hours to complete the assessment. This time frame suggests that, although testing experience allows for a comprehensive assessment, more time may be needed to perform all the proposed tasks.
The time variation among experts may reflect individual differences in work methodologies, familiarity with the content covered in the Guide, and perceived task complexity. These results provide valuable insight into optimizing the guide to make it more efficient and accessible without compromising the quality and comprehensiveness of the assessment performed by the experts.
The results from the experts regarding the planning, specification, and execution of the tests indicate that only the ESP3 expert covered all the topics of the Guide (**UG6**), and was the only one to automate the tests. This expert, who acts as Test Lead and has significant experience in Software, IoT, and Interoperability testing, demonstrated a comprehensive understanding and complete application of the Guide.
The other five experts (ESP1, ESP2, ESP4, ESP5, and ESP6) did not use all the topics in the Guide. They reported not using the tools provided in the "Suggested Tools" topic (**UG5**), which may indicate that the understanding was complex. Test automation, which requires advanced knowledge, may have been a challenge for these experts, contributing to the omission of this topic. In addition, the lack of familiarity with the recommended tools may have led to less confidence in applying automated tests.
Regarding the measures that experts used in their evaluation (**UG7**), **Table 20** illustrates the distribution and diversity of the measures collected by experts. The results show that three experts (ESP1, ESP3, and ESP6) collected measures during their evaluation, whereas the other three (ESP2, ESP4, and ESP5) did not. This finding suggests that experts with deeper experience and expertise in the three testing domains were more inclined to collect a broader set of metrics.

Table 20. Measurements used by the experts.

ID	Measurement
ESP1	Average response time for information requests; Success rate in integration with transportation systems (M05 and M06);
ESP3	Average response time for information requests; Success rate in integration with transportation systems; Level of compliance with communication protocols; Time required to adapt the application to new transportation systems; and Level of security of data transactions (M05, M06, M10, M11 and M12).
ESP6	Average response time for information requests and Time needed to adapt the application to new transportation systems (M05 and M13)

Question **UG9** seeks to identify whether the experts detected failures during their tests. **Table 21** gives an overview of the failures identified by experts during their evaluation of Rottas UFC.

Table 21. Failures identified by experts.

ID	Failure Description
ESP1 and ESP3	*The application experienced periods of inactivity that prevented access to routes and schedules.*
ESP1 and ESP3	*The application was slow to update information.*
ESP1 and ESP3	*The user interface is not very intuitive, making navigation difficult.*
ESP1 and ESP2	*Outages that affected the ability to access information about bus schedules and locations.*
ESP1 and ESP2	*Bus arrival time estimates were incorrect*
ESP1 and ESP2	*There were difficulties in integrating with an external app, affecting the updating of route and schedule information.*
ESP4 and ESP5	*It presented unexpected failures or crashes during use.*
ESP4 and ESP5	*Problems in communication between the application and the transportation systems resulted in outdated information.*

The results indicate that the experts identified failures in several Non-Functional Requirements (NFR), such as availability, *Interoperability*, performance, usability, and instability.
In addition, the six experts reported no difficulties in classifying and assigning severity to the identified failures (**UG10**) using the template provided in the instrumentation. The absence of such difficulties suggests that the Guide provides adequate tools and resources to facilitate the identification and analysis of problems in IoT applications. This result reinforces the perception that the Guide is intuitive and efficient, effectively guiding the experts during evaluation.

8.1.3 General Perception of the Use of the Interoperability Testing Guide

To evaluate perceptions of the Interoperability Testing Guide, we formulated 9 questions to determine whether the Guide is helpful to experts. These questions are presented in **Table 22**. Questions PG1-PG5 were designed using a Likert scale, while PG6-PG9 were open-ended.

Table 22. Questions about the Interoperability Testing Guide.

ID	Description
PG1	Did using the Interoperability Testing Guide facilitate test planning?
PG2	Did using the Interoperability Testing Guide guide you in creating test cases?
PG3	Was the Interoperability Testing Guide helpful in selecting measures?
PG4	Was the Interoperability Testing Guide helpful in calculating the cost-benefit?
PG5	Did the Interoperability Testing Guide give you a better understanding of Interoperability Testing?
PG6	Among the topics in the Guide, mark those you found most relevant for IoT Interoperability Testing.
PG7	Was there any topic in the Guide that you did not understand? Justify.
PG8	List the positive aspects of the Interoperability Testing Guide for IoT.
PG9	List the improvement points of the Interoperability Testing Guide for IoT.

The results indicate a trend of approval regarding the usefulness of the *Interoperability* Testing Guide among experts. In the first three questions (**PG1**, **PG2**, and **PG3**, all six experts (100%) strongly agreed with the statements that the Guide facilitated the planning of *Interoperability* tests, guided the creation of test cases, and was helpful in the selection of measures. These results indicate a positive and consistent perception of the Guide's effectiveness.
Regarding **PG4**, five specialists (83%) strongly agreed that the Guide helped calculate the cost-benefit of an *Interoperability* test. Only one participant (17%) was neutral, suggesting that while the overall perception is positive, there is a slight indication that the Guide could provide greater clarity or detail on this aspect.
In **PG5**, four participants (67%) strongly agreed, and two participants (33%) partially agreed that the Guide provided a better understanding of *Interoperability* testing. This result indicates that the Guide aids understanding of *Interoperability* testing, but it could still be refined to improve clarity.
We calculated the mode based on responses to the five Likert-scale questions. The overall mode for the perception of using the *Interoperability* Test Guide was 5. This result indicates that most experts evaluated the Guide's usefulness positively, highlighting its effectiveness in planning and executing IoT application tests. However, the results suggest further topics that could be refined to enhance the Guide's clarity and practical applicability.
Regarding the open-ended questions, **PG6** aims to identify the most relevant topics of the Interoperability Testing Guide based on the experts' evaluations. **Table 23** presents the topics they highlighted.

Table 23. Most relevant guide topics for *Interoperability* testing.

Guide Topic	Expert
Definition of Characteristics	ESP1
Interoperability Testing Challenges	ESP1 ESP2 ESP3 ESP4 ESP5 ESP6
Environment Configuration	ESP1 ESP3 ESP4 ESP5
Cost-Benefit	ESP1
Example of Guide Usage	ESP1 ESP3 ESP4 ESP5 ESP6
Characteristic Correlations	ESP2 ESP3
System Integration	ESP2 ESP3
Tool Suggestions	ESP2 ESP3 ESP4 ESP5 ESP6
Data Semantics	ESP5

The results reveal that the topic "Challenges of *Interoperability* Testing" was the most mentioned, with all six experts (100%) highlighting its importance. Indeed, this topic is crucial for identifying and addressing specific challenges in *Interoperability* testing in IoT.
The topics "Tool Suggestions" and "Example of Guide Use" were both highlighted by five experts. The high relevance of these topics indicates the need for practical

tools and clear examples for effective testing.

The topics "Test Environment Configuration" and "Correlation of Characteristics" were mentioned by four and two experts, respectively, reflecting the importance of properly preparing the testing environment and understanding correlations in *Interoperability* characteristics. The topics "Definition of Characteristic", "Cost-Benefit", and the subcharacteristic-related topics such as "Data Semantics" and "System Integration" were mentioned less frequently, suggesting that, while relevant, they are not considered priorities by all experts.

PG7 aims to identify whether any part of the Guide was unclear to the experts and to have them justify any such difficulties. The results reveal five main strengths: 1) Clarity and Didactic Approach; 2) Practical Usefulness; 3) Organization and Examples; 4) Ease of Understanding; and 5) Quality of the Material. **Table 24** presents the positive aspects highlighted by the experts.

Table 24. Positive Aspects of the IoT *Interoperability* Testing Guide.

ID	Expert's Comment
Clarity and Didactic Approach	*An educational support tool with objective definitions that greatly facilitates understanding for those unfamiliar with the field.*
Practical Usefulness	*The Guide is useful for several software testing activities, such as planning, specification, and execution of tests.*
Organization and Examples	*The Guide is very well organized, uses an easy-to-follow sequence, and the test case examples are beneficial.*
Ease of Understanding	*It greatly simplifies understanding and the steps required to conduct a thorough test.*
Quality of the Material	*This was a well-produced guide that allowed me to conduct robust testing based on the provided instructions.*

While the Guide's overall feedback was positive, a few areas for improvement were noted in question **PG9**. **Table 25** presents the suggestions offered by the experts who responded to this question.

The results show that the Guide is a valuable resource that combines clarity, organization, and practicality to support software testing activities. It is particularly effective for beginners, offering foundational knowledge and practical guidance. Incorporating targeted improvements—such as more detailed examples and streamlined sections—could enhance its utility and impact, becoming an indispensable tool for software testing professionals.

9 Discussion

This section presents key reflections on the research questions that guided the study, focusing on evaluating the Interoperability characteristic in IoT applications, the testing approaches used, and the associated challenges. The following subsections address each research question individually.

Table 25. Improvements for the *Interoperability* Testing Guide.

ID	Expert's Comment
"Example of Guide Use"	*It would be interesting to include an example/template on creating an interoperability test plan.*
"Tool Suggestions"	*I understand the intent to cover everything in a general context, but I would not include this section.*
"Correlation of Characteristics"	*I did not find much utility in the correlations section.*

9.1 RQ1. What aspects should be considered to evaluate the interoperability characteristic in IoT applications?

This RQ aims to identify the core aspects that should be considered when evaluating Interoperability in IoT systems. By addressing this question, we first analyzed existing testing approaches in the general software testing domain. The goal was to understand how Interoperability was commonly evaluated. Then, we focused on studies that aim to test Interoperability in the IoT domain[14]. In particular, the structure of the proposed guide is based on the findings of our previous work[Carvalho *et al.*, 2022].

This two-step analysis supported the definition of both the structure and content of the proposed testing guide. Evaluating *Interoperability* in IoT environments requires a systematic approach, considering the **diversity of protocols, standards, and technologies** involved [Rayes and Salam, 2019][Atzori *et al.*, 2010][Noura *et al.*, 2019]. Defining **clear testing objectives** and **selecting representative scenarios** are fundamental steps toward an effective evaluation.

Related studies, such as [Zaidi *et al.*, 2009], highlight the importance of integration tests that **simulate real-world scenarios** to evaluate the communication capabilities among heterogeneous devices. However, such tests often face challenges related to scalability and the robustness of the applications under test.

In this direction, Rath et al. [Rath *et al.*, 2018] recommend the adoption of **specialized tools** as a promising strategy, as they enable test automation and help identify compatibility issues. However, many of these tools lack a **well-defined methodological structure** that supports a comprehensive evaluation of Interoperability.

The literature reveals a lack of specific methodologies for evaluating Interoperability in IoT. In their absence, general software strategies such as integration and system testing are often adapted. However, these approaches do not adequately address the heterogeneity of devices, protocols, and environments inherent to IoT ecosystems.

[14] The complete list of selected studies is available at: `https://drive.google.com/file/d/1rKLYOSfSaAkwyg-9runtOTK1vSa2cBD9`

To address these gaps, we propose an Interoperability Testing Guide structured based on established practices and a well-defined methodology. This guide aims to support practitioners and researchers by providing clear procedures for conducting effective interoperability testing.
The importance of structured evaluation approaches is also reinforced by the findings of Carvalho et al. [Carvalho *et al.*, 2022] and Caldas et al. [Caldas, 2023], which highlight the need for structured and systematic evaluation approaches in IoT.

9.2 RQ2. How effective is the guide-based approach in evaluating the interoperability characteristic in IoT applications?

This research question focuses on evaluating the practical value of the proposed guide. The goal is to determine whether the guide's structure facilitates the testing process and whether its content is effective in assessing interoperability aspects, providing a systematic and comprehensive evaluation approach. Additionally, the evaluation aims to capture different user perspectives (*e.g.,* beginners and experts from academia and industry) regarding the guide's benefits, limitations, and potential adoption in their testing activities.
To assess the guide's effectiveness, three different evaluations have been conducted. The first used the TAM model to assess user acceptance of the *Interoperability Testing Guide*, covering five dimensions: *Perceived Usefulness*, *Ease of Use*, *Future Use Intention*, *Impact on Test Efficiency*, and *Overall Satisfaction*. The second was a controlled experiment involving two groups: one that used the guide and the other that did not. This evaluation enabled a practical assessment of the structure and content of the guide by utilizing a real-world IoT application. The third evaluation involved the analysis of experts and provided valuable feedback, which contributed to the refinement of the guide.
The results of all three evaluations were positive, demonstrating the effectiveness of the *Interoperability Testing Guide*. The TAM evaluation revealed a global mode of 5, indicating that "strongly agree" was the most recurrent response among participants for the majority of the questions, thus demonstrating strong acceptance and satisfaction with the Interoperability Testing Guide. The controlled experiment statistically demonstrated that the guide reduces the effort required to plan interoperability tests compared to traditional approaches and provides well-defined test case specifications that effectively support failure detection. Notably, the group using the guide identified more failures and detected all the failures found by the group without the guide. Finally, the evaluation by experts confirmed that the guide helps plan and execute interoperability tests. Additionally, the guide demonstrated the ability to support IoT failure detection through relevant test cases and measures.
Given the combination of user feedback, experimental results, and expert analysis, the results demonstrated substantial evidence of the guide's effectiveness in supporting interoperability testing in IoT applications. The guide enhances test planning and execution, facilitating the detection of interoperability failures through structured and well-defined test cases. These findings corroborate the guide-based approach as a practical and valuable tool for improving the quality of IoT systems.

10 Lessons Learned

The development of this research brought a series of challenges and lessons learned that were fundamental for creating and evaluating the interoperability testing guide for IoT applications. We organized these lessons into three main categories to facilitate reading and understanding: Methodology, Experiment Execution, and Tools and Automation.

- **Methodology**
  - **LL01 - Complexity in instantiating the approach**. The practical application of the guide highlighted the need to adapt the methodology proposed by [Carvalho *et al.*, 2022] to the specificities of different scenarios. It was challenging to clearly define the interoperability subcharacteristics in different contexts, which reinforced the importance of a flexible and modular approach in the guide. In addition, creating concrete and detailed examples was essential to help evaluators understand the application of the interoperability concept.
  - **LL02 - Technical understanding of the concept of interoperability**. Despite being a widely discussed topic in the literature, IoT interoperability still generates divergent interpretations among experts and less experienced users. Many participants had difficulty distinguishing between technical and semantic interoperability during the evaluation. This learning highlighted the need to include clear definitions and practical examples in the guide and the Wiki.
  - **LL03 - Complexity in conducting controlled experiments**. Running controlled experiments highlighted the importance of creating diverse and representative scenarios. The lack of real data and limited access to real test environments made validation more challenging. This learning reinforces the need to expand the scope of case studies in future work, exploring more complex and varied environments.
  - **LL04 - Balance between technical rigor and practical applicability**. One of the most significant challenges was finding the balance between the technical depth of the guide and its practical usability. Technical experts required more detail in the metrics and sub-features, while less experienced users preferred a simplified approach. This experience highlighted

the importance of building documentation that caters to different levels of technical knowledge.

- **Experiment Execution**

  * **LL05 - Difficulties in obtaining suitable IoT applications**. During the experiments, it was a great challenge to identify and access IoT applications that were sufficiently representative to validate the test guide. Many available systems were proprietary, had access restrictions, or did not have adequate technical documentation for analysis. This difficulty highlighted the need to build simulated scenarios to complement the experiments and explore partnerships with companies that develop IoT solutions.
  * **LL06 - Difficulty in recruiting experts for evaluation**. Identifying qualified experts with relevant experience in software testing and IoT was a great challenge. Despite significant efforts, time and resource constraints made it difficult to involve a larger number of evaluators. This challenge highlights the importance of building academic and industrial collaboration networks and planning longer periods for the evaluation phase.
  * **LL07 - Barriers to continuous feedback**. Although the feedback from experts was valuable, structuring the iterative stages of the evaluation was complex due to the limited availability of participants. Strategies had to be developed to consolidate the feedback received and prioritize the most critical improvements. This learning reinforces the importance of establishing clear processes for collecting and analyzing feedback in future studies.
  * **LL08 - Difficulties in collecting data for acceptance analysis**. The TAM (Technology Acceptance Model) application to assess the acceptance of the guide revealed that the questionnaires used could have been more detailed and targeted. Many participants provided generic responses, which limited the depth of the analysis. This point highlighted the need to invest more time in the design of data collection instruments to obtain more accurate insights.

- **Tools and Automation**

  * **LL09 - Challenges in automating the guide**. Creating the Wiki to support the guide was a significant advance. However, implementing automation revealed limitations, such as integrating complex functionalities, such as cost-benefit calculation. Although functional, the tool can still be expanded to include simulations and more detailed reports.

11 Threats to Validity

During the three evaluations of the *Interoperability* Testing Guide, some threats to validity were identified. These threats are categorized into internal and external validity, following the structure proposed by Wohlin et al. (2012) [Wohlin *et al.*, 2012]. Internal validity refers to the reliability and absence of bias in the evaluation, while external validity deals with the replicability of the evaluation. The identified threats and the mitigation strategies adopted are presented below.

11.1 Internal Validity

Threat 1: Lack of prior knowledge of participants

- **Description:** Some participants had no experience with *Interoperability* testing could affect the evaluation's quality.
- **Mitigation:** Introductory class discussions were held before the evaluations, and a questionnaire was applied beforehand to assess participants' knowledge.

Threat 2: Limited number of participants

- **Description:** The evaluation had only six participants for the TAM model and 12 for the controlled experiment, which could affect the representativeness of the results.
- **Mitigation:** Separate evaluations were conducted to explore different aspects of the guide: acceptance (TAM), practical use (controlled experiment), and technical validation with experts.

Threat 3: Limited time for the controlled experiment

- **Description:** The experiment was conducted in a single two-hour class, which may have restricted in-depth exploration of the guide.
- **Mitigation:** A detailed script and clear instructions were provided in advance to maximize the efficiency of the available time.

11.2 External Validity

Threat 4: Use of a simulated environment

- **Description:** The experiment used Rottas-UFC to simulate the operation of a bus, which may have impacted the applicability of the results to real environments.
- **Mitigation:** The tool was configured by the owner to realistically represent the vehicle's operation without altering the sequence of activities.

Threat 5: Evaluation with only one IoT application

- **Description:** The use of a single application may limit the generalizability of the results.
- **Mitigation:** An application that covers basic IoT functionalities was chosen, ensuring that *Interoperability* could be tested in a representative manner.

The mitigation strategies adopted aimed to minimize the identified threats and strengthen the reliability and replicability of the results of this study on the Interoperability Testing Guide in IoT applications.

12 Related Work

We conducted a literature review that explored approaches that address the gaps in *Interoperability* testing, focusing on the recent advancements in IoT application development, including studies, guidelines, techniques, methodologies, conceptual frameworks, and specific testing processes. The main goal was to understand current approaches and identify potential areas for advancement and improvement in the field of *Interoperability*.

12.1 General Approaches to Interoperability Testing in IoT Applications

Gunathilaka et al. [Gunathilaka *et al.*, 2016] proposed the *SoftGrid* system, a software-based smart grid testing platform designed to evaluate *interoperability* and security in IoT substation solutions. *SoftGrid* was developed to validate the accurate translation of messages sent and received through *gateways*, ensure proper routing of these messages, and confirm effective communication with Intelligent Electronic Devices (IEDs). In addition, it verifies the transmission and reception of messages between the IEDs and the control center, ensuring accurate processing and delivery. This system provides a systematic approach to assessing *interoperability* in substation environments, such as power transformation and distribution units, focusing on efficiency, performance, and security. However, its application is confined to substations and does not address broader *interoperability* concerns across diverse IoT applications.

Kuzniar et al. (2012) [Kuzniar *et al.*, 2012] introduced a technique for testing the *interoperability* of *OpenFlow switches*, essential components in software-defined networks. The study's main goal was to automate the identification of test inputs that produce inconsistent behaviors among different implementations of these switches. The method began with individual testing of each switch, applying a variety of inputs to observe and record their behavior. With the results, an analysis was conducted to identify whether a common set of inputs could trigger inconsistencies across implementations. This approach involved a detailed comparison of switch responses to identical inputs, enabling the detection of discrepancies that could undermine *interoperability* in software-defined networking environments. While effective, this technique is narrowly focused on *OpenFlow switches*. It does not provide a general framework for testing *interoperability* across other IoT applications—a key focus of the present study.

Mujjiga et al. (2007) [Mujjiga and Sukumaran, 2007] proposed a framework for the automated generation of *interoperability* tests based on the "Model Testing" technique. This process involves developing device-specific models aligned with manufacturer standards, described in the System Annotation Language (SAL). These models represent device behavior according to the Consumer Electronics Control (CEC) protocol, which aims to enhance functionality and *interoperability* within systems. The framework includes the manual creation of device models, followed by automated test generation using the *sal-atg* tool to evaluate compliance with the composite system model. While the approach enables the verification of device functionality against established standards, it lacks flexibility due to its reliance on predefined models and limited coverage of generated test cases. Additionally, the authors highlight challenges posed by optional features and vendor-specific attributes, suggesting an extension to the SAL language to improve adaptability. In contrast, the present study proposes a comprehensive *interoperability* testing guide focused on validation rather than model generation. Our guide encompasses 12 topics, including sub-characteristics, detailed properties, metrics, and correlations between *interoperability* and other IoT attributes.

12.2 Guidelines for Interoperability Evaluation in IoT: Guides, Standards, and Checklists

Carvalho, Lelli, and Andrade (2022) [Carvalho *et al.*, 2022] proposed a structured testing approach for evaluating IoT application characteristics, with a focus on performance. Our previous work introduced a performance evaluation guide organized into 11 topics, including the definition of performance characteristics, abstract test cases, metrics, tools, and more. The guide specifically addressed the "Performance" characteristic, subdivided into three subcharacteristics: "Temporal Behavior," "Resource Usage," and "Capacity." The guide was evaluated by experts and through a controlled experiment, demonstrating positive impacts on assessing IoT performance and identifying inherent flaws. In this work, we adapt the methodology to develop a test guide for *interoperability*, refining the original structure by adding a new topic that addresses testing challenges.

The ISO/IEC 21823:2022 standard [ISO/IEC 21823, 2022] provides guidelines to ensure *Interoperability* in IoT systems, facilitating effective communication and operation across various devices and systems. It defines *Interoperability* within IoT contexts, establishes specific requirements, and outlines methods and techniques for achieving it, including the adoption of common standards and protocols. The standard also introduces a *framework* to assess and improve *interoperability*, promoting the integration of multivendor devices. The benefits of the standard include simplified integration of IoT devices, adherence to uniform protocols, and improved operational efficiency.

Caldas (2023) [Caldas, 2023] proposed a *checklist* for evaluating *Interoperability* in IoT devices within smart home contexts. The checklist identifies *Interoperability* levels (technical, semantic, and organizational), defines a baseline of common devices and applications, and examines device components relevant to *Interoperability*. It evaluates aspects such as communication technologies, API availability, manufacturer documentation, and direct internet communication. Although practical for evaluating residential IoT devices, this method is limited to a predefined set of questions and does not address detailed sub-characteristics, properties, or metrics of *Interoperability*.

Motta et al., (2017) [Motta *et al*., 2017] explored the evolution of *Interoperability* in modern software systems, considering the influence of emerging technologies such as IoT, smart objects, and contextual systems through a literature review. They identified five dimensions of *Interoperability*: exchange capability, cooperation, integration, system relationship, and ownership, alongside seven key characteristics, including adaptive behavior, availability, compatibility, compliance, dynamic connection, and standardization. These dimensions provide a conceptual foundation for effective interaction in diverse systems.

Silva (2019) [Silva, 2019] introduced *ScenarIoT*, a checklist designed to evaluate IoT devices and system interactions across different environments and configurations. The checklist addresses key aspects such as component identification, functional and non-functional requirements, and device interactions. It categorizes and documents sensors, actuators, and tags while assessing characteristics like performance, security, privacy, reliability, scalability, and *Interoperability*. Additionally, it inspects data and command flows between devices to ensure efficient operation.

Unlike Caldas (2023) [Caldas, 2023], Motta et al., (2017) [Motta *et al*., 2017], and Silva et al., (2019) [Silva, 2019] we propose a test guide tailored to *Interoperability* testing, structured into 12 topics that include abstract test cases, properties, metrics, and correlations with other IoT characteristics, enabling comprehensive validation across varied scenarios.

12.3 Comparative Summary of the Studies

Based on the analysis of the aforementioned studies, distinct aspects and characteristics were identified in this research. **Table 26** groups the relevant studies, considering the flexibility of the methodologies (whether the methodology can be applied to different IoT contexts), the specific proposal (the main method or approach proposed by each study), and the testing phase that the study focuses on (planning, specification, execution, or others) addressed in each work. Moreover, the table allows the visualization of gaps and opportunities for improvement in existing approaches, highlighting areas where the current research can offer innovative and more effective contributions. Thus, a direct comparison can be established, illustrating how the methodology of this research advances compared to previous approaches.

We observed that this research shares similarities with the approaches adopted by other studies in testing *Interoperability* in IoT applications. However, some studies do not explicitly clarify the most effective approaches for this type of testing, especially regarding the flexibility of the methodologies and the phases of the testing process.

The distinctive aspect of this research lies in the construction of a structured guide in topics, which addresses the definition and contextualization of *Interoperability* in a comprehensive manner. This guide covers all phases of the testing process, including planning, specification, and execution. Additionally, the proposed guide includes the definition of test cases, suggestions for tools to automate the tests, abstract test cases, metrics, and use case examples, offering a complete structure that can be applied in various contexts and types of IoT applications.

13 Conclusion and Future Work

In this research, we proposed a guide-based approach for *Interoperability* Testing in IoT. The goal is to offer structured guidance for conducting interoperability testing. The *Interoperability* Testing Guide was developed to address the testing challenges in IoT, such as the complexity of the IoT architecture, the diversity of IoT protocols, and the heterogeneity of devices. Indeed, we cataloged 20 testing challenges concerning *Interoperability*.

The approach aims to address evident gaps, such as the lack of standardization in interoperability testing procedures, the absence of approaches to support this testing, and the lack of recommendations for suitable tools to assess *Interoperability* in IoT environments. We developed a guide based on two literature reviews to address these gaps. The first review was comprehensive, aiming to understand the *Interoperability* characteristic deeply. The second review focused on developing the specific topics of the guide.

To evaluate the *Interoperability* characteristic, we leverage ISO 30141 [ISO/IEC 30141, 2018] to divide this characteristic into four subcharacteristics: *Data Semantics, Communication Protocols, System Integration, and Network Protocols*. Thus, the testing guide is structured into 12 topics, covering essential aspects such as the interoperability subcharacteristics (definitions, contextualization, properties, test cases, and measurements)

We conducted three evaluations to assess the guide's effectiveness: (i) an evaluation using the TAM model (ii) a controlled experiment; and (iii) an evaluation with domain experts. Their results were positive, indicating that the guide is effective in testing planning and execution and in identifying IoT-specific failures.

In future work, several directions can be explored to expand this research. One direction is to apply the Interoperability testing guide in an industry context; case studies can then be used to evaluate its application across diverse scenarios and assess the guide's effectiveness in real-world environments.

Table 26. Related Work

Study	Flexibility	Proposal	Testing Phase	Characteristic
Zaid et al., (2009)	No	Contextual Signatures	Execution	Interoperability
Gunathilaka et al., (2016)	No	Smart Networks	Execution	Interoperability
Rath et al., (2016)	No	Automated Testing	Execution	Interoperability
Kuzniar et al., (2012)	No	OpenFlow Switches	Execution	Interoperability
Mujjiga et al., (2007)	No	Model Testing	Planning	Interoperability
Haj (2018)	No	Automated Testing	Execution	Interoperability
Motta et al., (2017)	No	Systematic Review	Planning	Interoperability
Caldas (2023)	Yes	Checklist	Verification	Interoperability
Silva et al., (2022)	Yes	Checklist	Specification	Interoperability
ISO/IEC 21823:2022	Yes	Guidelines for Interoperability	Evaluation	Interoperability
Carvalho et al., (2022)	Yes	Performance Test Guide	Planning, Specification, Execution	Performance
This research	Yes	Interoperability Testing Guide	Planning, Specification, Execution	Interoperability

Declarations

The authors would like to thank the Federal University of Ceará (UFC) for supporting this research. We also acknowledge GREat (Computer Networks, Software Engineering and Systems Research Group), a Research, Development and Innovation (R&D&I) center at UFC, for providing the research environment that enabled the development of this work. Special thanks are extended to Dr. Valéria Lelli Leitão Dantas for her guidance and supervision throughout this research. The authors also thank all students, professionals, and collaborators who participated in the evaluation activities and contributed to the discussions and improvements of this study.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author Contributions

Karina da Silva Castelo Branco conducted the literature review, developed the testing guide, performed the empirical studies, analyzed the results, and drafted the manuscript. Valéria Lelli Leitão Dantas supervised the research and contributed to the study design, analysis, and manuscript revision. Lavínia Matoso Freitas, Liana Mara Carvalho, and Paulo Ricardo Fernandes Rodrigues contributed to the research design, validation activities, analysis of results, and critical review of the manuscript. All authors read and approved the final version of the manuscript.

Competing Interests

The authors declare that they have no competing interests.

Availability of Data and Materials

The materials and resources supporting this study are publicly available.

- **The Interoperability Testing Guide is available at:** `https://github.com/karinascb/INTEROPERABILITY-TESTING-GUIDE-FOR-IoT`
- **The Wiki developed to support the guide is available at:** `https://laviniamatosof.github.io/interoperability_WIKI/`
- **The experimental materials used in the evaluation are available at:** `https://drive.google.com/drive/folders/1DozFXdxNVxTbI5fs3pI81TOuQ2JEclJU`

References

Almeida, R. L. A. (2018). *CHASE: checklist para avaliação da experiência do usuário em ambientes de internet das coisas*. Dissertação (mestrado em ciência da computação), Universidade Federal do Ceará, Fortaleza. Avaialable at:`http://repositorio.ufc.br/handle/riufc/57372`.

Andrade, S. and Luque, D. (2022). Interoperabilidade de sistemas aplicados às cidades inteligentes: Um estudo de mapeamento sistemático. In *Anais do X Workshop de Computação Aplicada em Governo Eletrônico*, pages 97–108. SBC. DOI: 10.5753/wcge.2022.222970.

Atzori, L., Iera, A., and Morabito, G. (2010). The internet of things: A survey. *Computer networks*, 54(15):2787–2805. DOI: 10.1016/j.comnet.2010.05.010.

Branco, K. C., Dantas, V., and Carvalho, L. (2024). Interoperability testing guide for the internet of things. In *Proceedings of the 30th Brazilian Symposium on Multimedia and the Web*, pages 188–196, Porto Alegre, RS, Brasil. SBC. DOI: 10.5753/webmedia.2024.242058.

Caldas, E. A. L. (2023). *Checklist para avaliação da interoperabilidade em dispositivos iot com foco em casas inteligentes*. Trabalho de conclusão de curso, Universidade Federal do Ceará, Campus de Quixadá, Quixadá. Available at:`http://repositorio.ufc.br/handle/riufc/73943`.

Cañas Betancur, D. C. and Hernández Sánchez, J. (2019). Comunicação assertiva em professores: diagnóstico e proposta educativa. *Praxis & Saber*, 10(24):143–165. DOI: 10.19053/22160159.v10.n25.2019.8936.

Carvalho, L. (2022). Wiki of the Performance Testing Guide for IoT Applications. Available at:`https://performance-testing-guide-staging.surge.sh`. Accessed on: April 29, 2025.

Carvalho, L. M., Lelli, V., and Andrade, R. M. (2022). Performance testing guide for iot applications. In *ICEIS (1)*, pages 667–678. INSTICC, SciTePress. DOI: 10.5220/0011090800003179.

Coutinho, A., Carneiro, E. O., and Greve, F. G. P. (2016). Computaçao em névoa: Conceitos, aplicações e desafios. *Minicursos do XXXIV SBRC*, pages 266–315. Available at:`https://www.researchgate.net/publication/309312669_Computacao_em_Nevoa_Conceitos_Aplicacoes_e_Desafios`.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3):319–340. DOI: 10.2307/249008.

dos Santos, M. C. (2016). Internet das coisas e sistemas inteligentes no jornalismo: o conceito de presença diluído entre as narrativas da complexidade urbana. *Comunicação & Inovação*, 17(34):21–39. DOI: 10.13037/ci.vol17n34.3769.

Ferreira, H. G. C. (2014). *Arquitetura de middleware para Internet das Coisas*. PhD thesis, University of Brasília Brasília, Brazil. Available at:`http://repositorio.unb.br/handle/10482/17251`.

Fienberg, S. E. (2006). When did bayesian inference become" bayesian"?. DOI: 10.1214/06-ba101.

Filho, F. G. S., Lelli, V., Santos, I. d. S., and Andrade, R. M. C. (2021). Correlations among software testability metrics. In *Proceedings of the XIX Brazilian Symposium on Software Quality*, SBQS '20, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3439961.3439972.

Ghasemi, A. and Zahediasl, S. (2012). Normality tests for statistical analysis: a guide for non-statisticians. *International Journal of Endocrinology and Metabolism*, 10(2):486–489. DOI: 10.5812/ijem.3505.

Giusto, D., Iera, A., Morabito, G., and Atzori, L. (2010). *The internet of things: 20th Tyrrhenian workshop on digital communications*. Springer Science & Business Media. DOI: 10.1007/978-1-4419-1674-7.

Gulla, J. A., Tomassen, S. L., and Strasunskas, D. (2006). Semantic interoperability in the norwegian petroleum industry. In *ISTA*, pages 81–93. Available at:`https://www.researchgate.net/publication/221141391_Semantic_Interoperability_in_the_Norwegian_Petroleum_Industry`.

Gunathilaka, P., Mashima, D., and Chen, B. (2016). Softgrid: A software-based smart grid testbed for evaluating substation cybersecurity solutions. In *Proceedings of the 2nd ACM Workshop on Cyber-Physical Systems Security and Privacy*, pages 113–124. DOI: 10.1145/2994487.2994494.

ISO 15926 (2011). Iso 15926 - industrial automation systems and integration - integration of life-cycle data for process plants including oil and gas production facilities. Available at:`https://www.iso.org/standard/50694.html`. Accessed on 28 Mar. 2023.

ISO/IEC 21823 (2022). Internet of Things (IoT) – Interoperability for IoT systems. (ISO/IEC 21823:2022). DOI: 10.3403/bsisoiec21823.

ISO/IEC 25000 (2005). ISO/IEC 25000:2005, software engineering - software product quality requirements and evaluation (SQuaRE). Available at:`https://www.bibsonomy.org/bibtex/2d23abc6b7a5745e618497de4cafcec97/reynares.e`.

ISO/IEC 25010 (2011). Systems and software engineering — systems and software quality requirements and evaluation (square) — system and software quality models. DOI: 10.3403/30215101.

ISO/IEC 25019 (2023). Systems and software engineering — systems and software quality requirements and evaluation (square) — quality-in-use model. DOI: 10.3403/30421230u.

ISO/IEC 30141 (2018). Information technology – Internet of Things (IoT) – Reference architecture. Available at:`https://www.iso.org/standard/65695.html`.

Kim, T. K. (2015). T test as a parametric statistic. *Korean Journal of Anesthesiology*, 68(6):540–546. DOI:

10.4097/kjae.2015.68.6.540.

Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., and Linkman, S. (2009). Systematic literature reviews in software engineering–a systematic literature review. *Information and software technology*, 51(1):7–15. DOI: 10.1016/j.infsof.2008.09.009.

Kuzniar, M., Peresini, P., Canini, M., Venzano, D., and Kostic, D. (2012). A soft way for openflow switch interoperability testing. In *Proceedings of the 8th international conference on Emerging networking experiments and technologies*, pages 265–276. DOI: 10.1145/2413176.2413207.

Legner, C. and Wende, K. (2006). Towards an excellence framework for business interoperability. In *19th Bled eConference eValues*, Bled, Slovenia. Available at:https://www.academia.edu/68299952/Towards_an_Excellence_Framework_for_Business_Interoperability.

Levene, H. (1960). Robust tests for equality of variances. pages 278–292. DOI: 10.2307/2285659.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22(140):1–55. Available at:https://psycnet.apa.org/record/1933-01885-001.

Maia, N., Macedo, G., Collins, E., and Dias-Neto, A. (2012). Aplicando testes ágeis com equipes distribuídas: Um relato de experiência. In *Anais do XI Simpósio Brasileiro de Qualidade de Software*, pages 365–372. SBC. DOI: 10.5753/sbqs.2012.15330.

Mann, H. B. and Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 18(1):50–60. DOI: 10.1214/aoms/1177730491.

Mattern, F. and Floerkemeier, C. (2010). From the internet of computers to the internet of things. In *From active data management to event-based systems and more*, pages 242–259. Springer. DOI: 10.1007/978-3-642-17226-$7_1$5.

Mattiello-Francisco, M. d. F. (2009). Inrob-uma abordagem para testes de interoperabilidade e de robustez de subsistemas de temporeal intensivos em software. *Doutorado em engenharia eletrônica e computação, Instituto de Tecnológico de Aeronáutica, São José dos Campos, SP*. Available at:https://bdtd.ibict.br/vufind/Record/ITA_426e3e1659698be580109a8707ca84a5.

Motta, R. C., De Oliveira, K. M., and Travassos, G. H. (2017). Rethinking interoperability in contemporary software systems. In *2017 IEEE/ACM Joint 5th International Workshop on Software Engineering for Systems-of-Systems and 11th Workshop on Distributed Software Development, Software Ecosystems and Systems-of-Systems (JSOS)*, pages 9–15. IEEE. DOI: 10.1109/jsos.2017.5.

Mujjiga, S. and Sukumaran, S. (2007). Modelling and test generation using sal for interoperability testing in consumer electronics. In *Proceedings of the second workshop on Automated formal methods*, pages 32–40. DOI: 10.1145/1345169.1345173.

Myers, G. J., Sandler, C., and Badgett, T. (2011). *The art of software testing*. John Wiley & Sons. DOI: 10.1002/9781119202486.

Noura, M., Atiquzzaman, M., and Gaedke, M. (2019). Interoperability in internet of things: Taxonomies and open challenges. *Mobile networks and applications*, 24:796–809. DOI: 10.1007/s11036-018-1089-9.

Nunes, P. R. d. A. F. (2011). *Validação de padrões de web services transacionais*. PhD thesis, Universidade de São Paulo. Avaialable at:https://www.teses.usp.br/teses/disponiveis/45/45134/tde-21072011-134559/publico/dissertacao_PauloNunesVersaoFinal.pdf.

Perera, C., Zaslavsky, A., Christen, P., and Georgakopoulos, D. (2014). Context aware computing for the internet of things: A survey. *IEEE Communications Surveys Tutorials*, 16(1):414–454. DOI: 10.1109/surv.2013.042313.00197.

Rath, F., Schemmel, D., and Wehrle, K. (2018). Interoperability-guided testing of quic implementations using symbolic execution. In *Proceedings of the Workshop on the Evolution, Performance, and Interoperability of QUIC*, pages 15–21. DOI: 10.1145/3284850.3284853.

Rayes, A. and Salam, S. (2019). *Internet of things from hype to reality: The road to digitization*. Springer. Book.

Rowland, C., Goodman, E., Charlier, M., Light, A., and Lui, A. (2015). *Designing Connected Products: UX for the Consumer Internet of Things*. O'Reilly Media, Inc., first edition. Book.

Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3-4):591–611. DOI: 10.1093/biomet/52.3-4.591.

Sicari, S., Rizzardi, A., Grieco, L. A., and Coen-Porisini, A. (2015). Security, privacy and trust in internet of things: The road ahead. *Computer networks*, 76:146–164. DOI: 10.1016/j.comnet.2014.11.008.

Silva, V. M. d. (2019). Scenariot: support for scenario specification of internet of things-based software systems. DOI: 10.5753/cbsoft$_e stendido$.2020.14628.

Soares, D. (2009). Interoperabilidade entre sistemas de informação na administração pública. *sl] Universidade do Minho*. Available at:https://repositorium.uminho.pt/entities/publication/1ca3cac8-103b-4f17-8430-ec27d084858a.

Venkatesh, V. and Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management science*, 46(2):186–204. DOI: 10.1287/mnsc.46.2.186.11926.

Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, pages 425–478. Available at:https://ssrn.com/abstract=3375136.

Vermesan, O. and Friess, P. (2011). *Internet of Things: Global technological and societal trends*. River Publishers. Book.

Weiss, L. C. *et al*. (2019). Interoperabilidade semântica: uma análise sob a perspectiva da abordagem ontológica de willard van orman quine. Available at:https://www.scielo.br/j/pci/a/XxxKxpcwLhc8yWD8kZhtGFk/?format=html&lang=pt.

Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., and Wesslén, A. (2012). *Experimentation in Software Engineering*. Springer Science & Business Media. DOI: 10.1007/978-1-4615-4625-2.

Zaidi, F., Bayse, E., and Cavalli, A. (2009). Network protocol interoperability testing based on contextual signatures and passive testing. In *Proceedings of the 2009 ACM symposium on Applied Computing*, pages 2–7. DOI: 10.1145/1529282.1529284.

Zanella, A., Bui, N., Castellani, A., Vangelista, L., and Zorzi, M. (2014). Internet of things for smart cities. *IEEE Internet of Things journal*, 1(1):22–32. DOI: 10.1109/jiot.2014.2306328.