MSA Strategies for End-of-Line Automated Test Equipment with Non-replicable Measurements

Eduardo Santos(✉), Eusébio Nunes, and Cristina Rodrigues

University of Minho, 4710-057 Braga, Portugal
ID11155@uminho.pt

Abstract. Measurement data quality is the fundamental pillar for statistical process control and reliable decision-making in modern industry. However, a persistent deficiency in scholarly literature regarding measurement system analysis (MSA) for complex electronic devices is observed. This article addresses MSA in automated industrial environments, focusing on automated test equipment (ATE) used in End-of-Line (EOL) testing in the automotive industry, where non-replicable measurements challenge traditional methods such as Gage R&R (GRR). The main objective is to propose and validate methodological strategies capable of evaluating the variability of the measurement system in these contexts, minimizing decision errors (Type I and II). To this end, an approach based on Expanded GRR combined with ANOVA was adopted, allowing the variance to be decomposed into different sources, such as parts, benches, nests, and repeatability error. The experimental study involved 19 devices tested in multiple configurations over 13 weeks, with statistical analysis performed using Minitab. The results revealed that the measurement system has a total GRR of 36.59%, considered inadequate, with significant contributions from repeatability (13.97%) and reproducibility (22.62%), in addition to statistically significant differences between benches, nests, and their interactions. It is concluded that classical methods are insufficient for non-replicable measurements, making it necessary to search for more robust models such as the Expanded GRR with ANOVA. The study reinforces the importance of considering reproducibility in ATE, improving equipment calibration and maintenance, and developing future research that relates measurement variability to product quality and industrial decisions.

Keyword: MSA · ATE · Non-replicable measurements

1 Introduction

In industries such as semiconductors and automotive electronics, Measurement System Analysis (MSA) [7] plays a critical role in quantifying measurement error and ensuring that variability introduced by the system does not compromise the correct functional classification of products. Inadequate characterization of the measurement system can lead to Type I errors - the improper rejection of conforming devices, directly impacting yield - or Type II errors - the improper acceptance nonconforming devices, with potential consequences for reliability and user safety [6].

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2027
O. Gervasi et al. (Eds.): ICCSA 2026, LNCS 16769, pp. 396–412, 2027.
https://doi.org/10.1007/978-3-032-30494-0_25

Concurrently, the quality of measurement data is widely recognized as a determining factor for the robustness of technical, economic, and strategic decisions [20, 27] gaining particular relevance in regulated sectors like the automotive industry. In this context, Integrated Device Manufacturers (IDM) are urged by customers and auditors to implement systematic and increasingly reliable methods for the classification of each Device Under Test (DUT), in compliance with regulatory and contractual requirements [3].

The centrality of metrological quality is reflected in international standards such as ISO 9001 [14], which requires the suitability of monitoring and measurement resources, and ISO 10012 [15], which establishes specific requirements for measurement management systems. Complementarily, ISO 5725 [13], the GUM (ISO/IEC Guide 98-3) [11], and ISO/IEC 17025 [17] provide the technical framework for evaluating precision, uncertainty, and validity of measurement methods and results. Together, these standards consolidate the requirement that measurement systems be "fit for purpose" and maintain their capability over time.

1.1 Classical MSA Approaches

MSA studies, particularly GRR, are based on the fundamental assumption of the replicability of measurements on the same object under identical conditions.

According to the Automotive Industry Action Group (AIAG) [7], the objective of MSA is to evaluate whether a measurement system consistently distinguishes conforming from non-conforming devices, thereby minimizing incorrect decisions associated with system variability. This methodology is widely applied in instrument validation, production process qualification, and structured quality improvement initiatives, allowing for the separation of inherent part-to-part variability from the variability introduced by the measurement system [4].

In mass production environments, the total error of the measurement system can be decomposed into components such as repeatability, reproducibility, number of distinct categories (ndc), bias, and temporal stability, as recommended by the AIAG manual [7]. Among these, repeatability and reproducibility - usually evaluated through GRR studies - are particularly relevant for the continuous improvement of process performance [6].

One of the main advantages of MSA, as established by the AIAG, lies in avoiding the complexity associated with full measurement uncertainty modeling [7]. Instead, it focuses on the practical evaluation of measurement system capability relative to the functional requirements of the process and product. In this sense, determining the capability of a measurement system is not only an essential element of quality and process improvement initiatives [4] but also a critical step that must precede any effort to improve the design or production process, as recommended by industrial best practices and standards [1].

It is important to note that despite its widespread use, a misconception persists in industrial practice that GRR studies directly evaluate measurement uncertainty. This interpretation is incorrect, as GRR is limited to analyzing the variability introduced by the measurement system and does not provide information regarding the interval within which the true value of the measurand is expected to lie, through calibration [22]. Thus, in accordance with ISO 10012 [15], ISO/IEC 17025 [17], and the principles of metrology established in the Guide to the Expression of Uncertainty in Measurement

(GUM), the calibration of measuring instruments must be viewed as a necessary and complementary activity to MSA, essential for ensuring metrological traceability and the proper evaluation of measurement uncertainty.

Classical MSA studies reveal significant limitations in industrial environments characterized by high automation, ATE, and complex hierarchical architectures.

1.2 MSA in Automatic Test Equipment

An ATE is a sophisticated hardware and software infrastructure designed to perform automatic measurements and diagnostics on components, sub-assemblies, or DUT. In the context of the automotive electronics industry, these systems are crucial for ensuring that products meet the rigorous performance and safety characteristics required by customers and quality auditors, such as those governed by the IATF 16949 [10] standard. In practice, the ATE replaces the human appraiser with a fully automated setup capable of performing parallel tests on multiple components more efficiently and at a lower cost. However, these systems introduce additional sources of variability, such as differences between various test sites and benches used [3].

The reliability of an ATE in the automotive electronics sector depends on the statistical equivalence of the different benches, meaning that different benches of the same ATE must produce identical results for the same device under test, regardless of the test bench, nest, time, or geographic location. The adequacy of this equipment is verified through MSA study methodologies, where the ATE is evaluated as an important component of the measurement chain to minimize Type I and Type II errors.

ATE introduces multiple sources of variability whose interaction frequently violates the assumptions of independence and homogeneity of variance inherent in standard MSA models [1]. This issue is exacerbated in non-replicable measurement processes.

This paper addresses this theme, building upon the study by Santos et al. [25], whose results highlighted the need for more in-depth MSA studies on ATE for automotive electronic systems. More specifically, this study aims to evaluate the measurement capability of an ATE used in End-of-Line (EOL) testing of automotive electronic devices, considering the differences between test benches and between nests within the same bench [3].

The remaining sections of this article are organized as follows: Sect. 2 presents EOL testing for automotive electronic devices, highlighting the framework and objectives of these tests, their main characteristics, and the non-replicable nature of many of the measurements obtained. Section 3 covers the main Experimental Design approaches in MSA studies, with special emphasis on those suitable for non-replicable measurements. Section 4 presents the methodology used in the MSA study described in detail in Sect. 5. Finally, Sect. 6 presents the main conclusions of this study.

2 EOL Testing in Automotive Electronic Devices

EOL testing in automotive electronic systems/devices comprises the set of tests performed at the final stage of the assembly process, immediately prior to camera measurement inspection (CMI), packaging, and shipping to the customer for vehicle integration.

EOL serves as the final technical barrier for verifying the functional, electrical, and, in many cases, communication compliance of automotive electronic modules, playing a critical role in ensuring the quality, safety, and reliability of the electronic device.

2.1 EOL Objectives and Tests Coverage

In the automotive industry, quality requirements are particularly demanding, both due to the regulatory framework (e.g., IATF 16949) [10] and the direct implications for road safety and brand reputation. Therefore, the primary objectives of EOL testing are:

- To verify the functional compliance of the final product.: activation of outputs (actuators, relays, drivers); reading of inputs (simulated sensors); validation of control algorithms; boundary condition response tests;
- To ensure that electrical and performance parameters are within specified tolerances: measurement of voltages, currents, and power consumption in different operating modes; verification of analog and digital signals; supply integrity and short-circuit protection tests;
- To confirm the correct programming, coding, and configuration of the software included in the final product;
- To guarantee the individual traceability of the tested product.

EOL tests can also include:

- Communication tests: verification of automotive protocols such as CAN, LIN, or Ethernet; validation of identifiers, firmware, and coding;
- Functional safety tests: in critical modules (e.g., braking systems), specific tests associated with functional safety requirements (ISO 26262) [16] may be performed.

Following these tests, the EOL system performs the device parameterization to standard values, preparing it for future pairing with the vehicle.

2.2 Main Characteristics of EOL Testing

EOL tests in automotive electronics present a set of distinctive characteristics, such as:

- High degree of automation between device and test system - with multiple channels and high hardware and software integration;
- Reduced cycle time – there is significant pressure to minimize test duration without compromising coverage;
- High functional coverage – unlike purely parametric tests, EOL tests tend to be highly functional, seeking to reproduce real operating scenarios;
- Full traceability – each tested device is typically associated with a unique serial number, time stamp, and a complete record of test results;
- Binary classification with strict criteria – EOL tests generally results in a pass/fail decision.

2.3 Non-replicable Measurements in Eletronic Devices

According to De Mast, J. and Trip, A. [18], a measurement is considered non-replicable when it violates one of two principles: (i) Temporal stability: the true value of the object changes over time; (ii) Robustness against measurement: the object is destroyed or significantly altered during the test.

Under these circumstances, the empirical separation between measurement system variability and the intrinsic variability of the object or measured characteristic becomes impractical, compromising the fundamental assumptions of conventional MSA methods. Despite the recognition of these limitations, a methodological gap persists regarding formally structured statistical frameworks for complex ATEs and non-replicable measurements [3].

Electronic circuits, due to their architecture, exhibit real variability, under controlled and identical conditions [29]. The variability can be classified at the hardware level, including manufacturing variations, thermal effects, electrical fluctuations, aging phenomena, and layout parasitic [21]. On hardware, variability derives from jitter and discrete components (i.e., resistors, capacitors, inductors) [23, 28]. Furthermore, manufacturing variations dictated by material composition and fabrication techniques also introduce variations [5]. Software also introduces variability: temporal (latency, task priority, interrupts); state (buffers, queues); environmental (drivers, libraries, OS versions, CPU load, and memory availability); as well as algorithmic and communication factors (bus latency, packet loss, and communication jitter). Other factors relate to the interaction between components, such as processing frequency, CPU load, and disk/memory read/write operations [19]. These are, in turn, associated with the system state (including switching between active and sleep modes of various sub-circuits) and interference from base software, such as operating system task scheduling-which includes instructions, routines, various software modules, and several network connections that also consume energy [8]. Both the circuit and ambient temperature also influence the results obtained [28].

In these cases, as in many other advanced industrial contexts, the fundamental assumption of classical MSA methods - the replicability of measurements on the same object under identical conditions - is not verifiable. Consequently, it is necessary to employ alternative MSA strategies that preserve statistical validity and metrological significance.

3 Experimental Designs in MSA Studies

The literature on MSA studies for non-replicable measurements suggests various methods or experimental strategies to isolate the different sources of variability within a measurement system, namely: Crossed designs [9, 22, 27]; Nested designs [9, 22, 27]; Expanded designs [22, 27]; Latin Square designs [26]; Mixed-Effects Models [1, 27]; and Type Gage 1 studies [22].

Resume. In non-replicable measurements, MSA requires a paradigmatic shift away from classical approaches. The choice of the most appropriate methodology depends on the system architecture, experimental constraints, and the objective of the analysis.

3.1 Crossed Design

This is the most common approach for evaluating measurement system variability through GRR studies. Every part is measured by every operator, and each operator-part combination is repeated several times, allowing for a fully crossed structure between factors. It allows the total variability to be decomposed into repeatability (equipment variation), reproducibility (variation between operators), and part-to-part variability. However, the disadvantages are (i) the need of a high number of measurements, which can involve significant cost and time; and (ii) assumption of stable experimental conditions and the absence of operator learning or fatigue effects. Several authors demonstrate usability for non-replicable measurements [9].

3.2 Nested Design

Nested designs constitute one of the most used approaches for non-replicable measurements. This model is essential when measurements are destructive or when parts cannot be measured more than once by the same operator. Each part is exclusive to one operator or condition, ensuring that two operators never measure the same device. However, the nested design cannot be used to estimate the reproducibility, and no additional factors can be added.

3.3 Expanded Design

Expanded designs are an extension of crossed and nested GRR designs, aimed at accommodating greater experimental complexity and it incorporates additional fixed or random factors (in addition to parts and operators), such as multiple pieces of equipment, environmental conditions, production batches, or operational levels, allowing for a more comprehensive assessment of measurement system performance in real-world contexts. Among the primary advantages of these experimental designs is their greater representativeness of the actual system, allowing for the capture of interactions between multiple sources of variability. However, their implementation is more complex, requiring more rigorous experimental planning, a larger volume of data, and more advanced statistical skills. Nevertheless, expanded designs do not, by themselves, solve the problem of the absence of replication.

3.4 Other MSA Designs

Latin Square designs provide an efficient, robust, and flexible method for controlling two sources of variability in non-replicable measurements, although they are restricted by equal factor levels and the assumption of no interactions [26]. Linear mixed-effects models offer a more versatile alternative by treating measurement system components as random effects to decompose variance in complex, unbalanced scenarios, despite requiring advanced statistical expertise [1, 27]. Type 1 Gage Study is also an alternative to classical MSA, for quick check of bias and repeatability of a single device through Cg and Cgk, used in ATE [22].

4 Methodology

Given the specific characteristics of ATE systems used in automated industrial environments - namely the high degree of automation and the lack of direct operator intervention - this MSA study adopted an approach based on the expanded GRR method. The implementation steps of this methodology are presented in the following sections.

4.1 Statistical Model

For data analysis, a random-effects ANOVA model was considered, allowing for the estimation of the various variance components (factors) associated with the measurement system. Establishing the factors to be included in the study is a fundamental aspect. Examples of factors included in the analysis are:

- Part (P): represents the intrinsic variability between the devices under test;
- Test Station (S): different test benches (ATE);
- Nest (N): physical contact interfaces with the bench (ATE);
- Repetition (ε): random error associated with the measurement process.

The general model may be expressed as:

$$Y_{ijkl} = \mu + P_i + S_j + N_k + \epsilon_{ijkl} \tag{1}$$

where:

Y_{ijkl} represents the observed value

μ is the global mean

P_i, S_j, N_k correspond to the random or fixed factors effect

ϵ_{ijkl} represents the repeatability error

The variance components are estimated based on this model, allowing for the quantification of the relative contribution of each source of variation.

4.2 Experimental Procedure

The experimental procedure was carried out in the following stages:

1. Selection of the variant based on statistical representativeness criteria;
2. Execution of the measurements in accordance with the established sampling plan;
3. Automatic data recording by the ATE.

4.3 Estimation of Variance Components

Total variance decomposition was performed based on Expanded and ANOVA analysis, allowing for the estimation of repeatability variance; variance associated with benches; variance associated with nests; bench-nest interaction variance; and part-to-part variance. Based on these components, the classical MSA indicators were calculated:

- Percentage of GRR (%GRR);
- Percentage of Repeatability;
- Percentage of Reproducibility;
- Number of Distinct Categories (ndc).

4.4 Acceptance Criteria

The suitability of the measurement system was evaluated based on criteria widely accepted in both literature and industrial practice:

- %GRR $< 10\%$: Suitable measurement system;
- $10\% \leq$ %GRR $\leq 30\%$: Acceptable system, depending on the importance of the application, the cost of the measurement device, or the cost of repair;
- %GRR $> 30\%$: Inadequate system; every effort should be made to improve it.

5 Case Study

This section presents the MSA study conducted on the ATE at the EOL assembly lines for automotive electronic products. These products are manufactured by a company in Portugal belonging to a multinational group. Currently, the company produces its devices according to defect acceptance criteria based on Class 3, with reference to the IPC Standard: Acceptability for Electronic Assemblies (IPC-A-610G) [12].

The analysis was processed through Minitab® Statistical Software.

5.1 ATE Characterization

ATE evaluates the device once it is fully assembled and encapsulated, in debug mode, under conditions as close as possible to real-world use, integrated at the end of the assembly/production line (see Fig. 1). The line has a total of eleven test benches with two nests each (two distinct physical interfaces). Each bench features a rack equipped with different measurement devices, allowing for the evaluation of a diverse set of functionalities and variables.

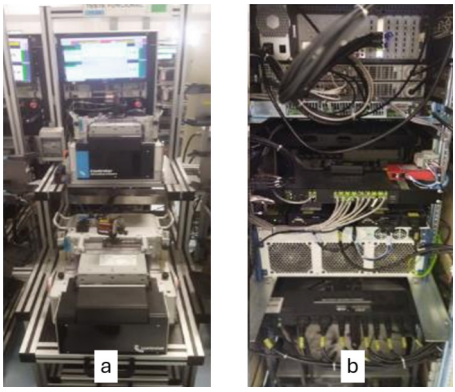

Fig. 1. Automatic EOL test bench, front (a) and rear view (b)

5.2 DUT Characterization

The product used in this study, hereafter referred to as the device (Fig. 2), is an infotainment system that combines informational functionalities (e.g., receiving signals from

the vehicle's peripherals) and entertainment (e.g., AM/FM, USB and Bluetooth protocols). Additionally, a group of Critical-to-Quality (CTQ) characteristics is established and measured, based on customer and supplier requirements.



Fig. 2. Electronic device tested in EOL

All products are functionally tested in ATE. Each functional characteristic corresponds to a specific number of numerical and attribute variables, with a final Pass/Fail. Furthermore, each type of measurement variable corresponds to a specific piece of equipment installed in the rack. As an example:

- For variables reflecting current consumption, the measurement equipment is a Digital Multimeter (DMM), common to both nests;
- For variables reflecting audio signal measurements or signal-to-noise ratios (FM, AM, DAB, etc.), the measurement equipment used is a Digital Signal Analyzer (DSA).

5.3 Adopted MSA Method in the Automotive Company

The method currently used by the automotive company to conduct MSA studies on ATE involves collecting a sample of ten consecutive parts of the same variant from the production process during a given shift.

The frequency of these studies is determined by customer requests or significant changes to the process or equipment. The statistical method employed is the Range method for a quick assessment and convenience.

The characteristics of each sample element are evaluated through four test runs on the same nest within one ATE. With the current MSA approach, the company assumes there are no differences between the independent ATEs and, consequently, considers reproducibility to be ignored. However, according to several authors [3, 24] significant differences may indeed exist between the ATEs. Furthermore, parts are collected sequentially (convenience sampling); therefore, the study fails to capture the real process variability. This highlights the need for a more detailed and methodologically demanding MSA study.

In this regard, the present paper acts as a continuation of the work by Santos, E. et al. [25], where results showed inconclusive trends and deviations between ATEs. That evaluation underscored the necessity for more in-depth MSA research, specifically applied to electronic systems.

5.4 Samples Preparation and Collection

On this device, true values change over time, therefore non-replicable measurements methods apply. The first step involves reducing measurement instability as much as possible. Mikulová, P et al. [20] address several general scenarios for this. Following the authors' recommendations, a homogeneous batch was selected - specifically, the same variant for the study - so that the devices could be considered "identical" for the analysis. Same variant was chosen, which consists of an electronic device with a set of identical attributes (same software version, hardware, and quantity and type of components), sharing the same unique identification name and ID number.

Sampling Plan. The study involved EOL testing on one of the production lines previously identified by Santos, E. et al. [25] as the most unstable. Sample collection took place over 13 non-consecutive weeks, and the respective tests were conducted throughout this same period. Benches with ID 1, 3, 5, and 7 were used for this study due to shift time constraints.

Since the line was not in production during the third shift (overnight), this time slot was utilized for the testing. An operator was assigned exclusively to perform the tests, representing a deliberate investment in extra time and cost.

The experimental plan was defined to ensure the representativity of process variability and the statistical robustness of the analysis. N devices were selected, assuring statistically the inclusion of some near the specification limits. A total of 19 devices for a given variant were subjected to the MSA test and were tested three times in each bench-nest combination. The sampling plan followed in this study is presented in the flowchart in Fig. 3.

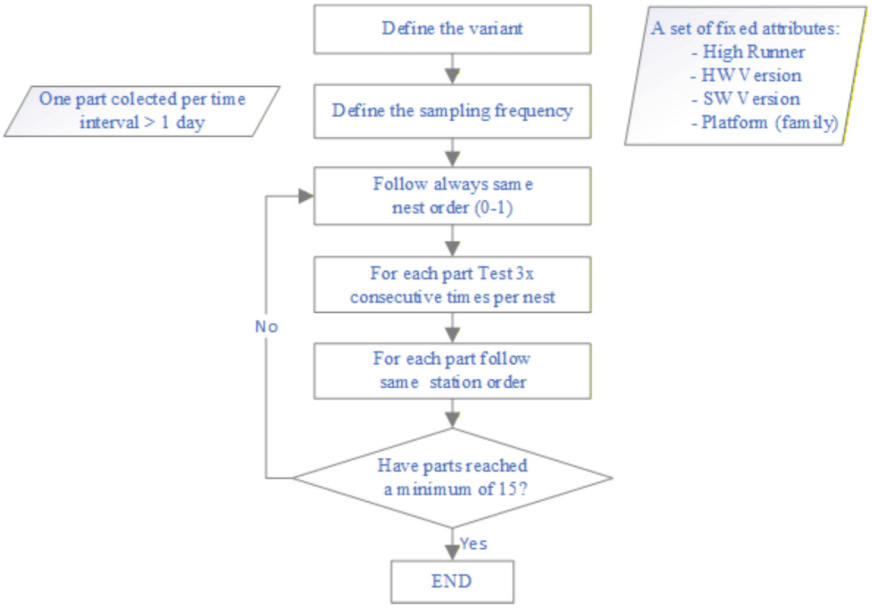

Fig. 3. Sampling plan

Part-to-Part Variation - Unique Parts Count. If the parts are stable and the randomness of the measurement error is homogeneous, a GRR can be performed [1].

Fourteen device variables were analyzed to select the most suitable ones for the study.

To ensure that the observed total variability was in greater proportion, due to actual part-to-part variation rather than random measurement error, variables that displayed a significant percentage of part-to-part variation were sought - specifically, those with a P-value equal to or less than 0.005 (very close to zero).

Additionally, the count of unique parts under this condition, for each variable, had to exceed 50% of the total tested samples. To achieve this, a Factorial ANOVA method (General Linear Model) was employed.

It was observed that five variables out of fourteen met this condition. Under these conditions, there is a high probability that repeatability and reproducibility represent a lower relative percentage than the part-to-part variation.

Normal Distribution. Before proceeding with the ANOVA calculations, the normality of the data for each selected variable was verified. This is a fundamental assumption for the validity of the GRR results, according to AIAG MSA [7].

Out of the five variables, it was found that only one, "102411 A2B+ Boost", fits a normal distribution at a 10% significance level, as shown in Fig. 4.

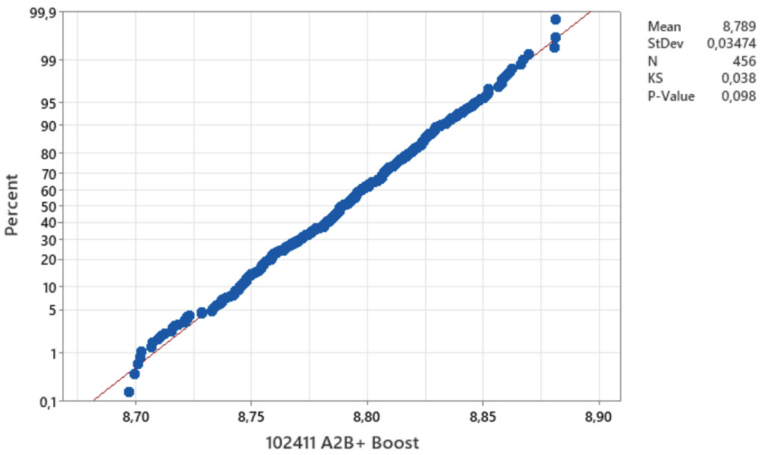

Fig. 4. Probability plot

It is worth noting that the CTQ customer variable, "Check Standby Current", follows a normal distribution; however, it exhibits very low part-to-part instability.

ANOVA – Time and Test Order Test. To determine whether significant differences exist in device measurements over time, a factorial ANOVA was performed. For the selected variable 102411 A2B + Boost, it was observed that significant differences occur over time between samples (measured week after week), and that significant differences exist within each sample according to the measurement sequence (test order).

Characterization of the Selected Variable. With the above criteria met, the variable selected for the study is "102411 A2B + Boost". Digital audio is measured using various parameters, specifically through the audio signal and current consumption.

This variable is one of the parameters used to measure digital audio; its function is to test the digital audio output via current consumption, measured in Volt. The signal is sent to the device ("unit" in the picture) through an individual generator installed in each nest, which then processes it and returns it to one of the common measurement devices on the bench, the DMM - as shown in Fig. 5.

The specification limits are LSL = 8.000000 V and USL = 9.000000 V.

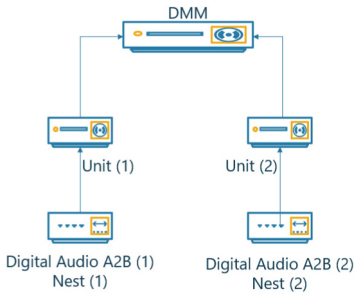

Fig. 5. Bench and Nest layout

5.5 MSA Expanded

Below, the application of the Expanded GRR and ANOVA method is highlighted. Since each ATE is independent and the nests share measurement equipment, it is considered an "operator".

5.6 Analysis Results

The ndc for this variable is equal to 1. According to the AIAG MSA manual [7], this indicates that the measurement system may be unable to identify distinct categories of parts.

However, according to VDA 5 [30], this parameter can be disregarded if the equipment resolution is at most 5% of the tolerance, by ensuring the quality of the readings. With adequate resolution, the determination of the ndc becomes dispensable, provided that the parts in the GRR have different measurements. Otherwise, the measurement system will fail to detect differences between parts [2]. In this study, the resolution of the equipment is 0.001% of the tolerance unit vector (0.1). Therefore, the ndc parameter can be safely ignored.

ANOVA study shows significant differences between devices ("Recoded Barcode"), test benches, nests and at interaction of Bench*Nest, as shown in Table 1.

Table 1. ANOVA Table with All Terms

Source	DF	Seq SS	Adj SS	Adj MS	F-Value	P-Value
Recoded Barcode	18	0.34503	0.34503	0.019168	320.99	0.000
Bench	3	0.08592	0.08592	0.028641	479.61	0.000
Nest	1	0.04280	0.04280	0.042797	716.67	0.000
Bench*Nest	3	0.04980	0.04980	0.016600	277.98	0.000
Repeatability	430	0.02568	0.02568	0.000060		
Total	455	0.54923				

α to remove interaction term = 0,05.

In the Components of Variation graph, %Contribution indicates the share of each source in the total variance and %Study Var shows the contribution of each source to the study's variability measured on the standard deviation scale. Both percentages sum to a cumulative 100%. See Fig. 6.

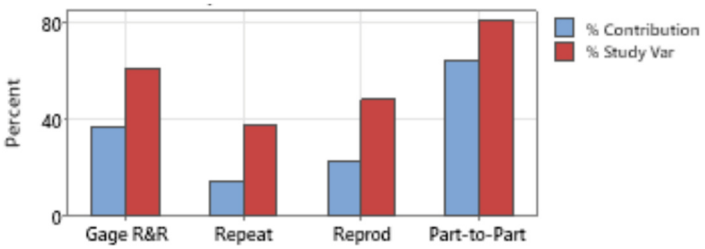

Fig. 6. Component of Variation

In Table 2, it is observed that the part-to-part component dominates, accounting for 63.41% of the variance between devices.

Total GRR accounts for 36.59% of the variance, resulting from the sum of Repeatability with 13.97% and Reproducibility with 22.62%.

Table 2. Variance Components result

Source	VarComp	%Contribution (of VarComp)
Total Gage R&R	0.0004566	36.59
Repeatability	0.0001743	13.97
Reproducibility	0.0002823	22.62
Bench	0.0001884	15.10
Nest	0.0000939	7.52

(*continued*)

Table 2. (*continued*)

Source	VarComp	%Contribution (of VarComp)
Part-To-Part	0.0007914	63.41
Recoded Barcode	0.0007914	63.41
Total Variation	0.0012480	100.00

The Sample Range chart, Fig. 7, represents the variability within each subgroup or part (max-min), where small and consistent ranges indicate good repeatability. This chart helps analyze and pinpoint specific causes of significant measurement system variation; it is observed that Bench 5 shows significant differences compared to the other benches. This suggests equipment instability on measurement capability.

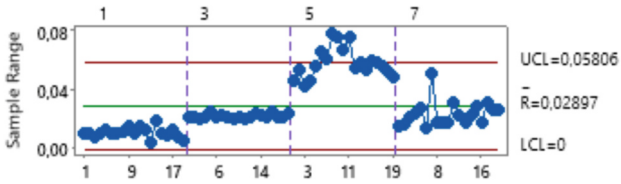

Fig. 7. R-Chart result

The Sample Mean chart, Fig. 8, shows the variability between different subgroups or parts and helps evaluate process trends (drift) and outliers that affect the measurement system assessment. A large part-to-part dispersion is observed (indicating a healthy difference between devices).

It is also observed that the process is out of control, regardless of the specification limits. The control limits are based on 3*standard deviation (STDEV = 0.0347); therefore, there is a significant probability of the population exceeding the specification limits.

In sum, the chart highlights a significant low systematic bias among three out of four benches (measuring similar levels) but may require verification of Bench seven (with higher average) and a review of the special causes for the points above the limits.

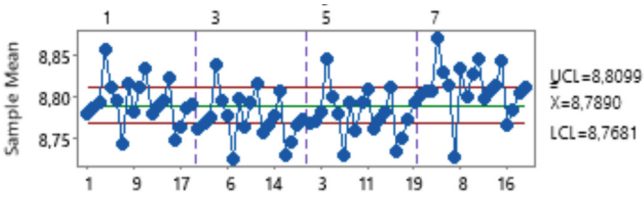

Fig. 8. X-Bar Chart result

In the Interaction Plot between Bench and Nest (Fig. 9), a significant difference is observed in the means within subgroups (max-min) - specifically between Nests - on Benches three and five.

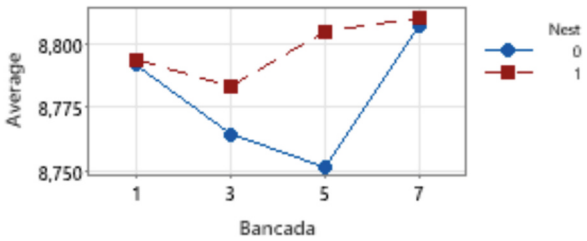

Fig. 9. Bench (*Bancada*)*Nest Interaction

6 Results Discussion

The results suggest that the measurement system shows inconsistencies which require attention and improvement. In benches three and five, disparity between nests must be analyzed as it raises suspicions regarding the consistency of the signal generators, or about the condition of the contact probes (test pins).

Out of the variables analyzed, only one fully met the criteria (normality and significant part-to-part variation). Nevertheless, it is worth mentioning that the conclusions impact an equipment (DDM) which is used for a significative number of different variables, in EOL.

Additionally, significant differences in averages are generally observed across all benches, particularly between benches one and seven. Since these two do not show internal variation between nests, it is recommended to investigate the causes of drift at the bench level and perform the necessary adjustments to the measurement equipment-in this case, the DMM (e.g., performing an immediate verification and reviewing the equipment's verification frequency).

Also, it is observed that variables representing CTQ characteristics from Original Equipment Manufacturers (OEM) customers may not always be suitable for GRR studies.

Finally, the MSA study confirms the hypothesis of significant differences between ATE and confirms that reproducibility must be incorporated in future MSA evaluations.

7 Conclusions and Future Work

To build upon the findings of this study, several research avenues are proposed to further bridge the gap between measurement system variability and industrial decision-making.

A longitudinal study is proposed to correlate MSA results with quality metrics, such as internal rejection rates and customer complaints. Identifying which specific measurement variables are the strongest predictors of field failures will allow for a more targeted allocation of maintenance resources.

It is necessary to continue optimizing "MSA capable" decisions, by evaluating which variables currently measured at the EOL are truly critical for the final decision-making process, including evaluation of current CTQ variables.

If validation of the ANOVA assumption is proven through rigorous ANOVA testing, it would simplify the calibration and cross-verification protocols across the entire production line.

Future research should evaluate the impact of data quality on pass/fail decisions, by quantifying how measurement uncertainty directly impacts the probability of Type I (false rejection) and Type II (false acceptance) errors. This involves modeling the "Gray Zone" around specification limits to determine the economic and quality risks associated with the current %GRR levels.

References

1. Awad, M., Erdmann, T.P., Shanshal, Y. Barth, B.: A measurement system analysis approach for hard-to-repeat events. Qual. Eng. **21**, 3, 300–305 (2009). https://doi.org/10.1080/08982110902852344
2. Beckert, S.F., Paim, W.S.: Critical analysis of the acceptance criteria used in measurement systems evaluation. Int. J. Metrol. Qual. Eng. **8** (2017). https://doi.org/10.1051/ijmqe/2017016
3. Bordogna, L., Brembilla, F., Pagani, A., Spinetta, M.: New R&R methodology in semiconductor manufacturing electrical testing. In: Proceedings - International Test Conference, pp. 410–419 (2022)
4. Burdick, R.K., Borror, C.M., Montgomery, D.C.: A review of methods for measurement systems capability analysis. J. Qual. Technol. **35**(4), 342–354 (2003). https://doi.org/10.1080/00224065.2003.11980232
5. Fledell, E., Grossman, B.: PCB manufacturing variation impact on high frequency measurement fixtures. In: 76th ARFTG Microwave Measurement Conference: Millimeter-Wave Measurements and Modeling, ARFTG 2010 (2010)
6. Flynn, M.J., Sarkani, S., Mazzuchi, T.A.: Regression analysis of automatic measurement systems. IEEE Trans. Instrument. Measure. **58**, 10, 3373–3379 (2009). https://doi.org/10.1109/TIM.2009.2025467
7. Automotive Industry Action Group. Measurement Systems Analysis (MSA) Reference Manual. Automotive Industry Action Group (2010)
8. Guo, C., Ci, S., Zhou, Y., Yang, Y.: A survey of energy consumption measurement in embedded systems. IEEE Access **9**(2021), 60516–60530 (2021). https://doi.org/10.1109/ACCESS.2021.3074070
9. Han, Y.-J., He, Z.: An applied study of destructive measurement system analysis. In: ICIEA 2007: 2007 Second IEEE Conference on Industrial Electronics and Applications, pp. 30–34 ((2007))
10. IATF 16949. Quality management system requirements for automotive production and relevant service parts organizations. IATF (2016)
11. International Organization for Standardization (ISO). ISO/IEC Guide 98-3:2008 Uncertainty of measurement — part 3: Guide to the expression of uncertainty in measurement (GUM:1995). International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC) (2008)
12. IPC-A-610G. Acceptability of Electronic Assemblies. IPC (2017)
13. ISO 5725. Accuracy (trueness and precision) of measurement methods and results — part 1: General principles and definitions. ISO (2023)

14. ISO 9001. Quality management systems — requirements. ISO (2015)
15. ISO 10012. Quality management — requirements for measurement management systems. ISO (2026)
16. ISO 26262. Road vehicles — functional safety (Series). ISO (2018)
17. ISO/IEC 17025. General requirements for the competence of testing and calibration laboratories. ISO (2017)
18. De Mast, J., Trip, A.: Gauge R&R studies for destructive measurements. J. Qual. Technol. **37**(1), 40–49 (2005). https://doi.org/10.1080/00224065.2005.11980299
19. Mazzola, S., Ara, G., Benz, T., Forsberg, B., Cucinotta, T., Benini, L.: 2025. data-driven power modeling and monitoring via hardware performance counter tracking. J. Syst. Architect. **167** (2025). https://doi.org/10.1016/j.sysarc.2025.103504
20. Mikulová, P., Plura, J., Knop, K.: Repeatibilty and reproducibility studies for non-replicable tests. Syst. Safety: Hum. – Techn. Facil. – Environ. **2020**, 275–284 (2020)
21. Raji, M., Tajary, A., Ghavami, B., Pedram, H., Zarandi, H.R.: Statistical leakage power optimization of asynchronous circuits considering process variations. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (2011), pp. 126–136 (2011)
22. Razumić, A., Runje, B., Alar, V., Štrbac, B., Trzun, Z.: A review of methods for assessing the quality of measurement systems and results. Appl. Sci. (Switzerland) **15**(17) (2025). https://doi.org/10.3390/app15179393
23. Roeder, J., Altmeyer, S., Grelck, C.: Can we trust our energy measurements? A study on the Odroid-XU4. arXiv e-prints (2022). arXiv:2206.10377. https://doi.org/10.48550/arXiv.2206.10377
24. Santos, E.: Development of a method to reduce the false positive rate in the equipment's final inspection for the automotive industry. Master Dissertation. University of Minho (2023)
25. Santos, E., Silva, S., Nunes, E., Rodrigues, C.: Non-Destructive Evaluation (NDE) Of functional characteristics of electronic devices for the automotive industry: the quality of data. In: ICOPEV - 6th International Conference on Production Economics and Project Evaluation, Braga (2024)
26. Shi, L., Ren, H., Wang, J., He, Z.: Latin square design for chip length machine vision measurement system analysis. Qual. Eng. **28**(4), 381–387 (2016). https://doi.org/10.1080/08982112.2016.1147049
27. Soares, W.D.O.S., Peruchi, R.S., Silva, R.A.V., Rotella Junior, P.: Gage R&R studies in measurement system analysis: a systematic literature review. Qual. Eng. **34**(3), 382–403 (2022). https://doi.org/10.1080/08982112.2022.2069505
28. Stanković, K., Vujisić, M., Kovačević, D., Osmokrović, P.: Statistical analysis of the characteristics of some basic mass-produced passive electrical circuits used in measurements. Measure.: J. Int. Measure. Confederat. **44**(9), 1713–1722 (2011). https://doi.org/10.1016/j.measurement.2011.07.007
29. Tofaily, S., Rais, I., Anshus, O.: Quantifying the variability of power and energy consumption for IoT edge nodes. In: Proceedings - 19th International Conference on Distributed Computing in Smart Systems and the Internet of Things, DCOSS-IoT 2023, pp. 577–584 (2023)
30. Verband der Automobilindustrie. VDA Volume 5: Measurement and Inspection Processes — Capability, Planning and Management. VDA QMC (Quality Management Center) (2021)