Tibuana
Journal of applied Industrial Engineering-University of PGRI Adi Buana
p-ISSN 2622-2027
DOI: https://doi.org/10.36456/tibuana.9.02.11482 *e*-ISSN 2622-2035

Evaluation of Measurement Instruments Using Fuzzy FMEA and MSA/GRR for Determining Calibration Intervals

Vinanda Asti Ningrum[1], Bambang Syairudin[2]
1,2 Department of Systems and Industrial Engineering, Faculty of Industrial Technology and Systems Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

e-mail : vinandaasti@gmail.com

***Abstract*— Laboratory testing instruments are critical for ensuring accurate measurement results and maintaining product quality. Failures or performance decline in these instruments may affect testing reliability. This study evaluates potential instrument failures using Fuzzy Failure Mode and Effects Analysis (Fuzzy FMEA) and assesses measurement system capability via Measurement System Analysis (MSA) with Gage Repeatability and Reproducibility (GRR). Five instruments, the Thickness Gauge, Roughness Tester, Tensile Tester, Density Tester, and Chromium Analyzer, were analyzed. Fuzzy FMEA results show the Roughness Tester has the highest Fuzzy Risk Priority Number (FRPN = 8.93), followed by the Chromium Analyzer (8.20), Tensile Tester (6.40), Thickness Gauge (5.47), and Density Tester (3.33). Historical data validate that higher FRPN values correspond to more frequent operational failures. GRR analysis indicates all instruments maintain acceptable measurement system performance (%GRR < 30%). Based on these findings, calibration intervals are recommended according to risk level and system stability, promoting efficient calibration management and reliable testing results.**

***Keywords*: Calibration Interval Evaluation, Fuzzy FMEA, GRR, Measurement Instruments, Measurement System Analysis**

I. INTRODUCTION

Measurement systems play a critical role in laboratories to ensure accurate measurement results and maintain consistent product quality, particularly in industries such as metal manufacturing, where adherence to technical specifications is essential. [1]–[3]. Accuracy, precision, and stability are key attributes of measurement systems, affecting the reliability of testing results and operational decisions. [4], [5]. Instrument failures or declines in performance may lead to false rejects or false accepts, which can compromise product quality, customer satisfaction, and operational costs.[6], [7].

In practice, most laboratories implement time-based calibration schedules, usually every 12 months, based on manufacturer recommendations or operational convenience. However, such schedules do not consider the actual condition of each instrument. [8], which can lead to over-calibration, wasting resources, or under-calibration, risking inaccurate results [9]. Conventional Failure Mode and Effects Analysis (FMEA) is commonly used to identify potential failures based on severity, occurrence, and detection, but it often suffers from subjectivity and ambiguity in Risk Priority Number (RPN) calculations. [10]–[12]

To address these limitations, Fuzzy FMEA integrates fuzzy logic, converting linguistic assessments (“low,” “medium,” “high”) into Triangular Fuzzy Numbers (TFN), resulting in a Fuzzy Risk Priority Number (FRPN) that better captures uncertainty and expert judgment variability. [13], [14]. Complementarily, Measurement System Analysis (MSA) using Gauge Repeatability and Reproducibility (GRR) quantifies variations caused by instruments (repeatability) and operators (reproducibility), providing a reliable evaluation of the measurement system’s performance. [13], [15], [16].

Fuzzy FMEA effectively prioritizes measurement risks by incorporating uncertainty and expert judgment, and MSA/GRR evaluates the capability and consistency of the measurement system. These approaches are rarely integrated into a unified framework for determining optimal calibration intervals. [11]. Calibration intervals are often established based on fixed schedules or

p-ISSN 2622-2027
DOI: https://doi.org/10.36456/tibuana.9.02.11482 *e*-ISSN 2622-2035

manufacturers' recommendations rather than the actual performance and risk profile of measurement instruments. [17], [18]. Consequently, instruments with stable performance may be calibrated more frequently than necessary, leading to unnecessary costs and downtime, whereas instruments with deteriorating performance may remain in service beyond appropriate intervals, increasing the risk of inaccurate measurements. [19].

By integrating Fuzzy FMEA and GRR, this study evaluates both the potential failures of instruments and the actual performance of measurement systems. The combined analysis serves as the basis for condition-based calibration intervals, adjusting calibration frequency according to each instrument’s risk level and stability. This approach allows laboratories to improve measurement reliability, optimize calibration resources, and reduce operational risks caused by inaccurate testing results [20], [21], [22].

II. RESEARCH METHODOLOGY

Research Approach

This study employed an applied quantitative research design to develop a risk-based calibration interval model for measurement instruments used in the metal manufacturing industry. The research aims to improve conventional time-based calibration practices by integrating instrument risk assessment with measurement system capability evaluation. The proposed framework combines Fuzzy Failure Mode and Effects Analysis (Fuzzy FMEA) and Measurement System Analysis (MSA) using Gauge Repeatability and Reproducibility (GRR). This integrated approach aligns with the principles of risk-based thinking promoted by ISO 9001:2015 and ISO/IEC 17025:2017, which emphasize identifying, evaluating, and controlling risks throughout the measurement process. Consequently, calibration intervals are determined according to both the operational risk and the actual performance of each measuring instrument rather than relying solely on fixed calibration schedules.

Research Object and Scope

The research was conducted in the quality control laboratory of a metal manufacturing company. Five critical measuring instruments routinely used in product quality inspection were selected as research objects:

1. Thickness Gauge
2. Surface Roughness Tester
3. Tensile and Elongation Tester
4. Density Tester
5. Chromium Content Analyzer

The study focuses on evaluating the reliability of these instruments through two complementary perspectives. First, the operational risks associated with each instrument are analyzed using Fuzzy FMEA to identify potential failure modes that may affect measurement reliability. Second, the capability of each measurement system is evaluated through MSA using Gauge Repeatability and Reproducibility (GRR). The integration of these analyses provides the basis for determining optimal calibration intervals.

Data Collection

This study utilized both primary and secondary data.

Primary data were collected through structured questionnaires administered to laboratory operators, maintenance engineers, quality assurance personnel, and laboratory supervisors. The questionnaire was designed to evaluate the three FMEA risk factors—Occurrence (O), Severity (S), and Detection (D)—using linguistic variables instead of numerical ratings. Secondary data consisted of:

1. historical calibration certificates,
2. maintenance records,
3. equipment failure reports,
4. measurement deviation records,
5. laboratory nonconformity reports,
6. Production quality records.

These secondary data were used to validate the identified failure modes and verify the consistency between expert assessments and actual instrument performance.

SOP Analysis and Operational Gap Identification

Before conducting the risk assessment, all Standard Operating Procedures (SOPs) related to instrument operation, calibration, maintenance, and verification were reviewed. The objective of this analysis was to identify discrepancies between documented procedures and actual laboratory practices. The identified operational gaps included:

1. inconsistent operating procedures,

p-ISSN 2622-2027
DOI: https://doi.org/10.36456/tibuana.9.02.11482 *e*-ISSN 2622-2035

2. inadequate preventive maintenance,
3. operator-related errors,
4. environmental influences,
5. delayed calibration activities.

The results of the SOP analysis became the primary reference for identifying potential failure modes during the Fuzzy FMEA process.

Instrument Risk Identification and Assessment Using Fuzzy FMEA

a. Failure Mode Identification

Potential failure modes for each measuring instrument were identified through:

1. SOP review,
2. direct field observations,
3. interviews with experts,
4. historical maintenance records,
5. calibration history,
6. Laboratory nonconformity reports.

Each failure mode was evaluated based on three classical FMEA criteria: Occurrence (O): likelihood of failure occurrence; Severity (S): impact of failure on measurement accuracy and product quality; Detection (D): ability of existing controls to detect the failure before affecting measurement results.

Linguistic Assessment

Instead of assigning crisp numerical scores, experts expressed their judgments using linguistic variables. Each linguistic term was represented by a **Triangular Fuzzy Number (TFN)**.

Linguistic Scale	**TFN (l, m, u)**
Very Low	(0,1,2)
Low	(1,2,3)
Medium	(2,3,4)
High	(3,4,5)
Very High	(4,5,6)
Critical	(5,6,7)

Aggregation of Expert Judgments

In Fuzzy FMEA, the three risk factors—Occurrence (O), Severity (S), and Detection (D)—are converted into Triangular Fuzzy Numbers (TFN) to handle the uncertainty in expert assessments:

$$\tilde{X} = (l, m, u)$$

The **Fuzzy Risk Priority Number (FRPN)** is calculated by combining the TFNs of the three factors:

$$\mathrm{FRPN} = \tilde{O} \times \tilde{S} \times \tilde{D}$$

This provides a representative risk value that accounts for the fuzziness in linguistic assessments such as "Low," "Medium," and "High"

Table 1. Linguistic Scale

Linguistic Scale	TFN (l, m, u)
Very Low / Rare	(0, 1, 2)
Low / Unlikely	(1, 2, 3)
Medium	(2, 3, 4)
High / Frequent	(3, 4, 5)
Very High / Very Frequent	(4, 5, 6)
Critical / Extremely High	(5, 6, 7)

Development of Risk Register and Risk Visualization

After the completion of the Fuzzy FMEA process, the next step was the preparation of a Risk Register to map potential failures for each instrument based on the analysis results. The Risk Register functions as a tool to classify and prioritize risks faced by each instrument according to the combination of Occurrence, Severity, and Detection values. These risk values were visualized in the form of a heatmap to facilitate the identification of areas with the highest risk levels. This heatmap assists in determining which instruments require further attention in terms of calibration and maintenance. The risk visualization also serves as the basis for decision-making in determining more adaptive and condition-based calibration intervals.

Evaluation of Measurement System Performance (MSA/GRR)

The evaluation of measurement system performance was conducted using Measurement System Analysis (MSA) with a focus on Gauge Repeatability and Reproducibility (GRR). GRR was used to measure two main types of variation in the measurement system: repeatability (variation caused by the measuring instrument) and reproducibility (variation caused by the operator). Data obtained from the GRR experiment was used to assess whether the measurement system utilized in the laboratory could provide consistent and reliable results. The GRR evaluation criteria applied were the standards commonly accepted in industry: if the

p-ISSN 2622-2027
DOI: https://doi.org/10.36456/tibuana.9.02.11482 *e*-ISSN 2622-2035

%GRR value is %, the measurement system is considered excellent; if 10% < %GRR ≤ 30%, the system is still acceptable with consideration; and if %GRR > 30%, the measurement system is considered inadequate.

The **GRR** method quantifies measurement system variation, separating repeatability (equipment) and reproducibility (operator) components. The total %GRR is calculated as:

$$\%GRR = \frac{\sqrt{\sigma^2_{\text{repeat}} + \sigma^2_{\text{reprod}}}}{\text{Tolerance}} \times 100\%$$

Interpretation:

- %GRR ≤ 10% → Excellent measurement system
- 10% < %GRR ≤ 30% → Acceptable
- %GRR > 30% → Inadequate

Integration of Fuzzy FMEA and GRR

The final stage integrated the results of Fuzzy FMEA and MSA/GRR to establish a risk-based calibration interval model. Each measuring instrument was classified according to two dimensions:

1. Operational risk represented by the FRPN value.
2. Measurement capability represented by the %GRR value.

The integration matrix provides the basis for calibration decision-making:

1. High FRPN + High %GRR: Calibration interval should be shortened, and corrective maintenance should be prioritized.
2. High FRPN + Low %GRR: Maintain calibration frequency while strengthening preventive maintenance and operational controls.
3. Low FRPN + High %GRR: Improve measurement capability before extending calibration intervals.
4. Low FRPN + Low %GRR: Calibration intervals may be safely extended, subject to compliance with organizational and regulatory requirements.

This integrated framework enables calibration decisions to be based on both operational risk and measurement performance, thereby improving laboratory reliability, reducing unnecessary calibration costs, and supporting continuous quality improvement.

III. RESULTS AND DISCUSSION

Historical Data Description

In this section, the analysis of historical data from testing instruments is essential for validating the results of risk evaluation conducted through Fuzzy FMEA and MSA/GRR. The historical data used include records of out-of-tolerance (OOT) incidents, drift deviations, downtime, and operator complaints that occurred during the testing period (January 2024 – February 2026). The analyzed instruments include the Thickness Gauge, Roughness Tester, Tensile Tester, Density Tester, and Chromium Analyzer.

Based on the collected historical data, instruments with high usage frequency, such as the Roughness Tester and Tensile Tester, experienced nonconformities or deviations more frequently than the other instruments. In contrast, instruments such as the Density Tester and Thickness Gauge demonstrated more stable performance with fewer nonconformity incidents.

Table 2. Historical Instrument Data (January 2024 – February 2026)

Instrument	OOT (Times)	Drift (Exceeded Limits)	Downtime (Hours)	Operator Complaints	Usage Frequency
Thickness Gauge	2	5	7	6	680
Roughness Tester	4	11	16	14	930
Tensile Tester	3	7	12	8	590
Density Tester	1	3	3	2	340
Chromium Analyzer	3	8	13	7	460

These data reveal that the Roughness Tester and Chromium Analyzer have higher numbers of OOT incidents, drift, and downtime, indicating that both instruments are more prone to nonconformities during the testing period. The Density Tester exhibited the most stable data, with minimal OOT incidents and downtime.

p-ISSN 2622-2027
DOI: https://doi.org/10.36456/tibuana.9.02.11482 *e*-ISSN 2622-2035

Fuzzy FMEA Results

Distribution of Occurrence–Severity–Detection Questionnaire Responses

This section presents the response patterns of respondents for the three main risk parameters in FMEA, namely Occurrence (O), Severity (S), and Detection (D). The analysis of this linguistic distribution aims to provide an initial overview of users' risk perceptions toward each laboratory instrument.

The distribution of linguistic values is important because it represents the initial stage in the fuzzification process, in which each linguistic category (e.g., Medium, High, Difficult) is converted into a Triangular Fuzzy Number (TFN) based on a predetermined fuzzy scale (see Appendix or Methodology Chapter).

This approach is consistent with recommendations in the Fuzzy FMEA literature (Tsai et al., 2017; Zadeh, 1975) as well as common practices in operator perception-based risk analysis according to ISO 31000:2018 and the metrological evaluation principles of ISO/IEC 17025:2017.

The questionnaire data consisted of 15 respondents for each instrument. The three risk parameters collected were:

1. Occurrence (frequency of potential failure occurrence)
2. Severity (severity level of failure impact)
3. Detection (ease of detecting potential failures)

Thickness Gauge

The following table presents the distribution of respondents' assessments for the Occurrence, Severity, and Detection parameters of the Thickness Gauge instrument.

Table 3. Thickness Gauge

Linguistic Scale	Occurrence	Severity	Detection
Very Rare	0	0	0
Rare	1	0	0
Medium	11	13	10
Frequent	3	0	0
Very Frequent	0	0	0
Low	-	1	-
High	-	1	-
Critical	-	0	-
Easy	-	-	0
Difficult	-	-	5
Very Difficult	-	-	0

The response distribution indicates that the majority of respondents assessed Occurrence as Medium, suggesting that the Thickness Gauge has a tendency to experience minor deviations periodically, but not at an alarming frequency. This is consistent with the characteristics of thickness measurement instruments, which experience wear and thermal drift over time.

The Severity assessment was also concentrated in the medium category. Thickness measurement deviations tend to have a moderate impact on product quality, especially for materials with tight tolerances. However, most errors can still be corrected through re-verification, so the "High" category was selected by only one respondent.

For the Detection parameter, there were two major groups: Medium (10 respondents) and Difficult (5 respondents). This indicates that some deviations, particularly micro drift, are not immediately visible to operators during daily measurements. The inability to detect small drift is consistent with findings in metrology literature, which state that micro deviations in thickness sensors are often only detected during calibration or standard verification.

The Thickness Gauge demonstrates a moderate risk level, with the dominant factors being difficult in detecting drift and moderate deviation frequency.

Roughness Tester

The following table presents the distribution of respondents' assessments for the Occurrence, Severity, and Detection parameters of the Roughness Tester instrument.

Tabel 4. Roughness Tester

Linguistic Scale	Occurrence	Severity	Detection
Medium	2	0	0
Frequent	11	0	0
Very Frequent	2	0	0
High	-	11	-
Critical	-	4	-
Difficult	-	-	11
Very Difficult	-	-	4

The Roughness Tester exhibited the highest risk pattern among all reviewed instruments. For the Occurrence parameter, responses were concentrated in the Frequent and Very Frequent

p-ISSN 2622-2027
DOI: https://doi.org/10.36456/tibuana.9.02.11482 *e*-ISSN 2622-2035

categories, indicating that performance deviations occur relatively often. This is consistent with the mechanical characteristics of stylus systems, which are sensitive to wear, contamination, and vibration, as reported in tribology and surface metrology literature. [23][24], [25].
The Severity parameter was dominated by the High and Critical categories. Errors in surface roughness measurements may result in false acceptance or false rejection of products, thereby significantly impacting production quality.
For the Detection parameter, most respondents assessed it as Difficult or Very Difficult, which is consistent with the characteristics of stylus-based sensor instruments, where damage or drift is often hidden and only detected when deviations become significant. ISO 4287 literature also emphasizes that sensor noise and stylus damage are common causes of deviations that are difficult to diagnose.
The Roughness Tester is considered a very high-risk instrument and, therefore, should become a priority in risk control and Fuzzy FMEA evaluation.

Tensile Tester
The following table presents the distribution of respondents' assessments for the Occurrence, Severity, and Detection parameters of the Tensile Tester instrument.

Tabel 5. Tensile Tester

Linguistic Scale	Occurrence	Severity	Detection
Rare	1	0	0
Medium	14	2	8
High	-	12	-
Critical	-	1	-
Difficult	-	-	7

The Tensile Tester demonstrated a medium-to-high risk pattern. The Occurrence parameter tended to fall within the medium category, reflecting failures related to jig alignment, load cell drift, or sample preparation errors commonly encountered in material testing.
For the Severity parameter, most respondents selected the High category, with one respondent assigning a Critical rating. This indicates that errors in tensile strength and elongation measurements have serious consequences for evaluating material integrity, as emphasized in ASTM E8/E8M standards.
The Detection distribution was divided between Medium and Difficult. Errors in the Tensile Tester may be detected through observation of stress–strain curves; however, load cell drift or internal damage is not always directly visible, making detection relatively challenging according to some respondents.
The Tensile Tester is categorized as a medium-to-high risk instrument, influenced by the significant impact of errors and the moderate difficulty of detection.

Density Tester
The following table presents the distribution of respondents' assessments for the Occurrence, Severity, and Detection parameters of the Density Tester instru ment.

Tabel 6. Density Tester

Linguistic Scale	Occurrence	Severity	Detection
Rare	11	-	-
Medium	4	11	4
Low	-	4	-
Easy	-	-	11

The Density Tester exhibited a low-risk pattern. Most respondents rated Occurrence as Rare, consistent with the instrument's stable construction and relatively simple measurement process.
The Severity parameter was distributed between Low and Medium, indicating that density measurement errors do not have major consequences in quality control processes, particularly since density values can still be verified through repeated measurements.
The Detection parameter showed very positive results, with most respondents selecting Easy. This indicates that performance errors in the Density Tester are usually clearly visible, such as reading fluctuations or unstable values.
In conclusion, the Density Tester is a low-risk instrument and therefore occupies a lower priority level in the risk mitigation plan.

Chromium Analyzer
The following table presents the distribution of respondents' assessments for the Occurrence, Severity, and Detection parameters of the Chromium Analyzer instrument.

p-ISSN 2622-2027
DOI: https://doi.org/10.36456/tibuana.9.02.11482 *e*-ISSN 2622-2035

Tabel 7. Chromium Analyzer

Linguistic Scale	Occurrence	Severity	Detection
Rare	1	0	0
Medium	14	0	0
High	-	14	-
Critical	-	1	-
Difficult	-	-	14
Very Difficult	-	-	1

The Chromium Analyzer demonstrated high-risk characteristics. The Occurrence parameter fell within the medium category, while Severity showed an extreme pattern in the High and Critical categories. This indicates that errors in chromium concentration measurements may have significant implications, especially regarding environmental regulations and metal finishing product quality compliance.

The Detection parameter was also categorized as Difficult and Very Difficult. According to optical instrument literature (Skoog & Holler, 2013), spectral noise, reduced light source intensity, and optical contamination often cause deviations that are not immediately detected by operators.

In conclusion, the Chromium Analyzer is a high-risk instrument, mainly due to the substantial impact of measurement errors and the difficulty in detecting early failure symptoms.

This questionnaire distribution analysis provides an initial overview of risk perceptions for each instrument. The results serve as the basis for the fuzzification process, in which all linguistic values are converted into Triangular Fuzzy Numbers (TFN). The fuzzification process and subsequent fuzzy calculations are presented in the following subsection (4.2.2).

Measurement System Analysis (MSA)

Gauge Repeatability and Reproducibility (GRR) was used to evaluate variations in the measurement system caused by the instrument itself (repeatability) and by operators (reproducibility). The GRR evaluation results indicate that instruments with higher %GRR values tend to exhibit greater variation in measurement performance between operators and instruments.

Table 8. GRR Evaluation Results for Instruments

Instrument	%GRR	Interpretation
Roughness Tester	9.57%	Excellent
Tensile Tester	7.31%	Excellent
Chromium Analyzer	19.33%	Very Good
Density Tester	8.41%	Excellent
Thickness Gauge	19.67%	Very Good

From the GRR results, instruments such as the Roughness Tester showed higher %GRR values, indicating that variation between operators and instruments is relatively significant. This demonstrates that calibration and operator training are highly important to ensure measurement consistency for these instruments.

The Density Tester and Thickness Gauge showed better performance with lower %GRR values, indicating that measurement variations in these instruments are more controlled.

Evaluation and Determination of Calibration Intervals

Based on the results of Fuzzy FMEA and GRR, calibration intervals are recommended to be adjusted according to the actual condition of each instrument. Instruments with high risk and high %GRR values, such as the Roughness Tester, should be calibrated more frequently, while instruments with low risk and low %GRR values, such as the Density Tester, may have longer calibration intervals.

Table 9. Recommended Calibration Intervals Based on Fuzzy FMEA and GRR

Instrument	Recommended Calibration Interval
Roughness Tester	Every 3 months
Tensile Tester	Every 6 months
Chromium Analyzer	Every 6 months
Density Tester	Every 12 months
Thickness Gauge	Every 9 months

These calibration interval recommendations indicate that calibration based on actual instrument conditions provides benefits in optimizing resources and improving the reliability of measurement results. More efficient calibration interval determination can reduce the risk of operational failures without compromising

p-ISSN 2622-2027
DOI: https://doi.org/10.36456/tibuana.9.02.11482 *e*-ISSN 2622-2035

product quality.

Final Fuzzy Risk Priority Number (FRPN) Ranking

The defuzzified Fuzzy Risk Priority Numbers (FRPN) for all measurement instruments are summarized in Table 4.3. This ranking illustrates the priority of instruments based on their potential failure risk, guiding calibration and maintenance strategies. Instruments with higher FRPN values require closer attention and more frequent calibration.

Table 10. Final Fuzzy Risk Priority Number (FRPN) Ranking

Instrument	FRPN (Crisp)	Risk Level	Ranking
Roughness Tester	8.93	Very High	1
Chromium Analyzer	8.20	High	2
Tensile Tester	6.40	Medium-High	3
Thickness Gauge	5.47	Medium	4
Density Tester	3.33	Low	5

This table shows the defuzzified FRPN values for each measurement instrument, indicating their risk levels and ranking. Higher FRPN values correspond to higher potential failure risk, which guides priority for calibration and maintenance. The FRPN values were obtained using the centroid method from Triangular Fuzzy Numbers (TFNs) of the Occurrence, Severity, and Detection ratings.

IV. Discussion

Risk Analysis Using Fuzzy FMEA

The results of the Fuzzy FMEA indicate that each measurement instrument has a different level of risk, which affects the recommendations for calibration intervals. Based on the assessment of Occurrence, Severity, and Detection, high-risk instruments such as the Roughness Tester and Chromium Analyzer require special attention in terms of calibration and routine maintenance.

The Roughness Tester showed the highest Fuzzy Risk Priority Number (FRPN), indicating that this instrument has a very high level of risk (Pacana & Olearczyk, 2017). This instrument frequently experiences nonconformities that are difficult to detect and have significant impacts on product quality. The difficulty in detecting errors before testing is completed is one of the reasons why the Roughness Tester should have a shorter calibration interval, namely every three months. This finding is consistent with previous studies showing that instruments with high FRPN values require greater attention in terms of maintenance and more frequent calibration. [26], [27].

In contrast, instruments such as the Density Tester and Thickness Gauge demonstrated lower risk levels, with significantly smaller FRPN values. The Density Tester exhibited low risk due to the rare occurrence of failures and low severity, where density measurement errors do not significantly affect final product quality. This finding also supports that low-risk instruments may be calibrated at longer intervals without affecting product quality. [28].

Evaluation of Measurement System Performance Using GRR

The results of the Measurement System Analysis (MSA) using Gauge Repeatability and Reproducibility (GRR) indicate that the %GRR value for the Roughness Tester is relatively high, meaning that measurement variation is more influenced by the instrument and operator than by the characteristics of the tested samples. This indicates that more frequent calibration is necessary to maintain the accuracy of measurement results and reduce the influence of human factors. [29]. This finding is consistent, which suggests that instruments with high %GRR values require routine calibration to maintain consistent measurement quality. [28].

Instruments with low %GRR values, such as the Density Tester and Thickness Gauge, demonstrated better measurement stability. With low %GRR values, the occurring variations remain acceptable within the specified tolerance limits, and these instruments do not require overly frequent calibration. This finding supports studies stating that instruments with stable performance may have longer calibration intervals. [30], [31]

Implications of the Research Findings

This study provides an important contribution to the management of laboratory instrument calibration through a more flexible and efficient condition-based approach. The Roughness Tester,

p-ISSN 2622-2027
DOI: https://doi.org/10.36456/tibuana.9.02.11482 *e*-ISSN 2622-2035

which demonstrated high risk and large measurement variation, requires greater attention in terms of calibration and maintenance to ensure consistent measurement quality. [29]. On the other hand, the Density Tester and Thickness Gauge, which demonstrated more stable performance, may be calibrated at longer intervals, thereby optimizing laboratory resource utilization and calibration costs. This risk-based approach is more efficient than the time-based approach commonly implemented in many laboratories. [30], [32].

The simultaneous use of Fuzzy FMEA and GRR provides significant advantages in identifying instruments that require more frequent calibration based on their risk levels and performance stability. By adopting this approach, laboratories can improve the reliability of measurement systems and extend instrument service life without compromising testing accuracy [32], [33].

Based on the results of this analysis, it is recommended that the Roughness Tester be calibrated every three months, while the Density Tester may have a longer calibration interval of every 12 months. The Thickness Gauge, which has a low % GRR and moderate risk level, may be calibrated every 9 months. This risk-based approach will not only improve calibration efficiency but also reduce costs associated with excessive calibration for instruments that do not require it. This is consistent with the findings of [34], [35], which suggests that condition-based calibration can reduce operational costs and improve measurement accuracy.

V. Conclusion

This study successfully developed a risk-based calibration system model for measurement instruments used in testing laboratories. Based on the results of the Fuzzy FMEA and GRR analyses, it can be concluded that high-risk instruments, such as the Roughness Tester, require shorter calibration intervals, namely every 3 months. This is because the Roughness Tester exhibited a high Fuzzy Risk Priority Number (FRPN), which reflects a high frequency of failures, significant impacts on product quality, and difficulties in detecting instrument errors. In contrast, instruments with low %GRR values and low risk levels, such as the Density Tester and Thickness Gauge, demonstrated better measurement stability and may be calibrated at longer intervals, namely every 12 months and 9 months, respectively, in accordance with the findings obtained.

The condition-based calibration approach proposed in this study enables more efficient calibration management, reduces excessive calibration for instruments that do not require it, and still maintains measurement accuracy for high-risk instruments. By using Fuzzy FMEA and GRR simultaneously, laboratories can improve efficiency in managing calibration resources, optimize calibration costs, and ensure product quality without compromising the reliability of measuring instruments.

Overall, this study provides an important contribution to improving the efficiency of calibration systems in testing laboratories, while also introducing a more data-driven model that aligns with the guidelines of ISO 9001:2015 and ISO/IEC 17025:2017 regarding risk management in calibration and quality control of measuring instruments. The recommendation from this study is to implement a more flexible risk-based calibration approach, with calibration intervals adjusted according to the actual performance and condition of the measurement instruments used.

References

[1] V. K. V. Meti, C. Soori, A. Talli, A. C. Giriyapur, P. Revankar, and I. G. Siddhalingeshwar, “A Project-Based Learning Approach to Measurement Systems Laboratory for Undergraduate Students,” *J. Eng. Educ. Transform.*, vol. 37, no. 4, 2024, doi: 10.16920/jeet/2024/v37i4/24157.

[2] K. Nishio *et al.*, “A digital laboratory with a modular measurement system and standardized data format,” *Digit. Discov.*, vol. 4, no. 7, 2025, doi: 10.1039/d4dd00326h.

[3] T. Kabir, “Digital Twin–Enabled Optimization of Electrical, Instrumentation, And Control Architectures In Smart Manufacturing And Utility-Scale Systems,” *Int. J. Sci. Interdiscip. Res.*, vol. 6, no. 1, pp. 404–451, 2025.

[4] K. M. T. A. Anick and T. Kabir, “Reliability-Centered Maintenance of Electrical Power and Control Systems Using Manufacturing-Based Asset Management and Quality Models,” *Am. J. Adv. Technol. Eng. Solut.*, vol. 2, no. 03,

p-ISSN 2622-2027
DOI: https://doi.org/10.36456/tibuana.9.02.11482 *e*-ISSN 2622-2035

pp. 29–59, 2022.
[5] N. V Martyushev, B. V Malozyomov, S. N. Sorokova, E. A. Efremenkov, D. V Valuev, and M. Qi, “Review models and methods for determining and predicting the reliability of technical systems and transport,” *Mathematics*, vol. 11, no. 15, p. 3317, 2023.
[6] A. Jabłonowski, A. Krasiński, and P. Kanty, “Laboratory tests of measurement systems in pile shafts,” *Soils Found.*, vol. 65, no. 4, 2025, doi: 10.1016/j.sandf.2025.101663.
[7] L. H. Lubis, D. A. Kasma, and M. I. Nasution, “Design of Measurement System for Laboratory-Scale Resistivity Meter with ADS1115 and Buffer,” *POSITRON*, vol. 15, no. 2, 2025, doi: 10.26418/positron.v15i2.99679.
[8] M. Berangi *et al.*, “Gradient boosting decision trees to study laboratory and field performance in pavement management,” *Comput. Civ. Infrastruct. Eng.*, vol. 40, no. 1, 2025, doi: 10.1111/mice. 13322.
[9] B. Boche, S. Temam, and O. Kebede, “Inventory management performance for laboratory commodities and their challenges in public health facilities of Gambella Regional State, Ethiopia: A mixed cross-sectional study,” *Heliyon*, vol. 8, no. 11, 2022, doi: 10.1016/j.heliyon.2022.e11357.
[10] J. Qin, Y. Xi, and W. Pedrycz, “Failure mode and effects analysis (FMEA) for risk assessment based on interval type-2 fuzzy evidential reasoning method,” *Appl. Soft Comput. J.*, vol. 89, 2020, doi: 10.1016/j.asoc.2020.106134.
[11] L. Ouyang, W. Zheng, Y. Zhu, and X. Zhou, “An interval probability-based FMEA model for risk assessment: A real-world case,” *Qual. Reliab. Eng. Int.*, vol. 36, no. 1, 2020, doi: 10.1002/qre. 2563.
[12] K. H. Chang, Y. C. Chang, and P. T. Lai, “Applying the concept of exponential approach to enhance the assessment capability of FMEA,” *J. Intell. Manuf.*, vol. 25, no. 6, 2014, doi: 10.1007/s10845-013-0747-9.
[13] W. Qiao, Z. Chen, and T. Hou, “Failure Mode and Effects Analysis of Accessibility in Frigid Zone Campus Buildings,” *Sustain.* , vol. 16, no. 8, 2024, doi: 10.3390/su16083437.
[14] I. Perić, K. Klarić, A. Pirc Barčić, K. Vukman, T. Sedlar, and P. Grošelj, “Optimising Risk Management in Wood-based Manufacturing: A Fuzzy AHP-FMEA Framework Approach,” *Bioresour.* , vol. 20, no. 2, 2025, doi: 10.15376/biores.20.2.2979-3001.
[15] J. Arumun, V. Ekanem, and E. Ashigwuike, “Review of gauge repeatability and reproducibility for dimensional measurements,” *Sch. Bull.*, vol. 10, no. 7, 2024.
[16] A. Razumić, B. Runje, V. Alar, B. Štrbac, and Z. Trzun, “A review of methods for assessing the quality of measurement systems and results,” *Appl. Sci.*, vol. 15, no. 17, p. 9393, 2025.
[17] E. A. Haque, “Design and Simulation of Sustainable Calibration Systems for Future Industrial Engineering Applications,” *Am. J. Adv. Technol. Eng. Solut.*, vol. 4, no. 03, pp. 60–99, 2024.
[18] M. D. Holloway, *Accuracy and Calibration Failures in Science and Engineering*. John Wiley & Sons, 2025.
[19] R. K. Mobley, *Maintenance fundamentals*. Elsevier, 2011.
[20] G. Geschwind, M. Vignal, M. D. Caballero, and H. J. Lewandowski, “Evidence for validity and reliability of a research-based assessment instrument on measurement uncertainty,” *Phys. Rev. Phys. Educ. Res.*, vol. 20, no. 2, 2024, doi: 10.1103/PhysRevPhysEducRes. 20.020125.
[21] S. Djenadic, M. Tanasijevic, P. Jovancic, D. Ignjatovic, D. Petrovic, and U. Bugaric, “Risk evaluation: Brief review and innovation model based on fuzzy logic and MCDM,” *Mathematics*, vol. 10, no. 5, p. 811, 2022.
[22] E. A. Haque, “The Role of Calibration Engineering in Strengthening Reliability of US Advanced Manufacturing Systems Through Artificial Intelligence,” *Rev. Appl. Sci. Technol.*, vol. 4, no. 02, pp. 820–851, 2025.
[23] R. M. Almutairi and M. Chikkanna, “Comparative Assessment of Testing Systems for Evaluation of Properties of

p-ISSN 2622-2027
DOI: https://doi.org/10.36456/tibuana.9.02.11482 *e*-ISSN 2622-2035

Different Newer Restorative Resins - An In Vitro Study," *Contemp. Clin. Dent.*, vol. 16, no. 4, 2025, doi: 10.4103/ccd.ccd_499_25.

[24] P. Podulka, C. Felho, I. Sztankovics, and L. Knapčíková, "Roughness evaluation of machined Ti-6Al-4V alloys with a study of surface topography measurement noise," *Meas. J. Int. Meas. Confed.*, vol. 257, 2026, doi: 10.1016/j.measurement.2025.118686.

[25] S. Fu, W. S. Kor, F. Cheng, and L. K. Seah, "In-situ measurement of surface roughness using chromatic confocal sensor," in *Procedia CIRP*, 2020, vol. 94. doi: 10.1016/j.procir.2020.09.133.

[26] Sherin Ramadhania, Dinanjar Husni Mubarok, Mislan Mislan, Asep Rahmatullah, Bambang Setyo Panulisan, and Dedy Khaerudin, "Quality Control Of Capsule A Products Using Statistical Process Control (SPC) And Fuzzy Failure Mode And Effect Analysis (F-FMEA) Methods At PT.LP," *Uranus J. Ilm. Tek. Elektro, Sains dan Inform.*, vol. 2, no. 1, 2024, doi: 10.61132/uranus.v2i1.80.

[27] V. Karanović, B. O. Ceylan, and M. Jocanović, "Reliable ships: A fuzzy FMEA-based risk analysis on four-ram type hydraulic steering system," *Ocean Eng.*, vol. 314, 2024, doi: 10.1016/j.oceaneng.2024.119758.

[28] M. Grennan and R. J. Town, "Regulating innovation with uncertain quality: information, risk, and access in medical devices," *Am. Econ. Rev.*, vol. 110, no. 1, pp. 120–161, 2020.

[29] G. Baek, C. Cha, and J.-H. Han, "AI chatbots for psychological health for health professionals: Scoping review," *JMIR Hum. Factors*, vol. 12, no. 1, p. e67682, 2025.

[30] Y. Deng, S. Du, D. Wang, Y. Shao, and D. Huang, "A calibration-based hybrid transfer learning framework for RUL prediction of rolling bearing across different machines," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–15, 2023.

[31] S. S. Shukla, R. K. Pandey, B. Gidwani, and G. Kalyani, "Calibration methods and procedures," in *Pharmaceutical Calibration, Validation and Qualification: A Comprehensive Approach*, Springer, 2023, pp. 29–40.

[32] H. Fereidouni, A. S. Hafid, D. Makrakis, and Y. Baseri, "F-RBA: A federated learning-based framework for risk-based authentication," *arXiv Prepr. arXiv2412.12324*, 2024.

[33] S. Roy, B. Mandal, and B. K. Das, "Quality assurance in laboratory practices and equipment maintenance: Ensuring precision, reliability, and compliance," in *Laboratory Techniques for Fish Disease Diagnosis*, Springer, 2025, pp. 79–103.

[34] I. Turkin, V. Leznovskyi, A. Zelenkov, A. Nabizade, L. Volobuieva, and V. Turkina, "The use of IoT for determination of time and frequency vibration characteristics of industrial equipment for condition-based maintenance," *Computation*, vol. 11, no. 9, p. 177, 2023.

[35] A. Martins, I. Fonseca, J. T. Farinha, J. Reis, and A. J. M. Cardoso, "Online monitoring of sensor calibration status to support condition-based maintenance," *Sensors*, vol. 23, no. 5, p. 2402, 2023.