AEROSPACE STANDARD

AS13003

Issued 2015-02

Measurement Systems Analysis Requirements for the Aero Engine Supply Chain

RATIONALE

The aerospace industry is heavily reliant on inspection to ensure that parts and assemblies delivered to the purchaser meet drawing requirements. There are many differing requirements across the aero engine supply chain, therefore, this standard is intended to harmonize these requirements into a single approach.

The determination of what needs to be inspected is covered in a separate standard AS13002.

This standard defines the essential requirements to establish acceptable measurement systems (for variable and attribute features) for use on aero engine parts and assemblies.

TABLE OF CONTENTS

TO PLACE A DOCUMENT ORDER: Tel: 877-606-7323 (inside USA and Canada)
Tel: +1 724-776-4970 (outside USA)
Fax: 724-776-0790
Email: CustomerService@sae.org
SAE WEB ADDRESS: http://www.sae.org

SAE values your input. To provide feedback on this Technical Report, please visit http://www.sae.org/technical/standards/AS13003

1. SCOPE

This standard defines the minimum requirements for conducting Measurement Systems Analysis (MSA) for variable and attribute assessment on characteristics as defined on the drawing or specification. It does not define the detailed analytical methods for each type of study as these can be found in existing published texts (see Section 2 for guidance).

1.1 Purpose

The aerospace industry is highly reliant on inspection to ensure that parts and assemblies meet drawing requirements. Ensuring that measurement systems are capable and repeatable is vital to the effectiveness of the inspection process.

The purpose of this document is to define the application of appropriate measurement system analysis tools and the acceptance criteria to be applied by the Aero Engine Manufacturers Supply Chain. It shall also provide guidance on the efficiency of application (read across) and mitigation strategies for non-capable measurement systems.

This standard defines the MINIMUM acceptance limits for measurement systems analysis. The purchaser may require different acceptance standards for specific applications.

There may be situations where alternative measurement systems analysis needs to be deployed. These should be agreed between the supplier and the specific purchaser prior to approval.

Case studies are included to provide practical examples of the application of these methods and further reading is also provided in Section 2.

2. REFERENCES

2.1 Applicable Documents

The latest issue of SAE publications shall apply. Nothing in this document shall supersede applicable laws and regulations unless a specific exemption has been obtained. The documents listed below are intended to support the requirements of this document and provide guidance on conducting MSA studies.

2.1.1 SAE Publications

Available from SAE International, 400 Commonwealth Drive, Warrendale, PA 15096-0001, Tel: 877-606-7323 (inside USA and Canada) or 724-776-4970 (outside USA), www.sae.org.

2.1.2 Automotive Industry Action Group (2010). Measurement System Analysis, 4th ed., Detroit, MI.

2.1.3 ASTM E2782, Standard Guide for Measurement Systems Analysis.

2.1.4 Wheeler, D. J. and Lyday, R. W., Evaluating the Measurement Process, SPC Press, Inc., Knoxville, TN, 2006.

2.1.5 "Evaluating the Measurement Process III - Using Imperfect Data" by Donald Wheeler.

2.1.6 Minitab support documentation – www.minitab.com

2.2 Definitions

ACCURACY: How close the measured value is to the 'accepted reference value'.

ACCURACY ERROR OR BIAS: The difference between the observed average value of measurements and a known true value or accepted reference value. Accuracy Error is measurement error not captured in most MSA evaluations and must be considered in the overall assessment of the measurement system. Bias is discussed further in 8.2.7 of this standard.

NOTE: Observed Accuracy Error may be additive to MSA error and the combined values shall not violate Table 2 requirements. An example of Accuracy Error would be when inspecting a feature with a known value of 0.006 the average observed value is 0.0075. This is an accuracy error of 0.0015.

This is not typically seen in off the shelf inspection devices but very common in purpose built inspection devices.

ACCURACY RATIO: The ratio between the total part tolerance and the total calibration tolerance of the measurement equipment.

ARTEFACT: An object of known size, shape, chemical composition, etc., that is used to test the measurement system condition by comparison of the known artefact characteristic against the measured result. Artefacts can be a component which is measured in a number of ways to establish an acknowledged known measurement value. Artefacts can also be calibration masters or an item which has been calibrated to accurately determine its characteristics.

ATTRIBUTE: A qualitative measure of a property that is of interest. This may be binary (pass/fail, good/bad, etc.) or ordinal if the values can be ranked (e.g., low, medium, and high).

CORRELATION: The degree to which two or more factors are statistically related to each other and vary together. This does not necessarily prove a cause and effect relationship just that they appear to show a link.

CRITICAL FEATURES: Those characteristics of an item which, if nonconforming, may result in hazardous or unsafe conditions for personnel using, maintaining or depending on the product; or which may prevent or seriously affect the satisfactory operation or functioning of the product. Specific Purchaser Feature Classifications may exist.

GAUGE REPEATABILITY & REPRODUCIBILITY (GR&R): A study used to determine how much variation is present in a measurement system by combining the equipment variation (repeatability) with the system variation (reproducibility). See 8.2.4 of this standard

LINEARITY: (1) An assessment of accuracy through the defined range of expected measurements in any inspection system. (2) The ability of the measurement system to consistently measure across the intended range with very little or no variability. Lack of Linearity is typically seen in gauges which inspect parts that change geometrically and materially at the same time, with instruments at fixed locations and angles. It is also typical for process related evaluations where process parameters may change throughout the evaluation cycle.

MAJOR FEATURES: Those characteristics of an item, other than critical, which if nonconforming, may result in operational or functional failure of the item, or which materially reduce the usability, physical or functional interchangeability or durability of the product for its intended purpose. Specific feature specifications will be defined by the purchaser.

MEASUREMENT SYSTEMS ANALYSIS (MSA): The process of evaluating the fitness for purpose of a measurement system, including methods such as Gauge R&R, Attribute Agreement, Bias assessment, Stability assessment, Linearity assessment, etc.

MINOR FEATURES: Those characteristics of an item which, if nonconforming, do not materially reduce the usability, physical or functional interchangeability or durability of the product, or are departure from established standards having no significant bearing on the effective use or operation of the product. Specific feature specifications will be defined by the purchaser.

NUMBER OF DISTINCT CATEGORIES: The number of distinct categories within your measurement data that the measurement system can distinguish.

PRECISION: How close the repeated measurement results are to each other.

PURCHASER: The purchaser issuing the purchase order for the part that is subject to the measurement capability acceptance.

REPEATABILITY: The ability of the measurement system to give the same result when measuring the same feature multiple times using the same elements of the system; e.g., gauge, operator, environment, fixture, etc.

REPRODUCIBILITY: The ability of the measurement system to give the same result when elements of the system, such as operator or environment, are changed.

RESOLUTION: The ability of the gauge to detect changes in the characteristic being measured and discriminate between measurement values.

STABILITY: A measure of how variation changes over time. This can be classified as short term or long term stability depending on the time frame involved.

SUPPLIER: A supplier is any manufacturer of systems, sub-systems, assemblies, components and materials for use within the Aero Engine Supply Chain.

VARIABLE: Something that is liable to change and is not a fixed value. There may be one or more causes of the variation either acting independently or together. Variable or continuous data may take on any value within a finite or infinite interval depending on the resolution of the measurement system used to capture that value.

3. APPLICABILITY

This standard is intended for businesses that design and/or manufacture products throughout the Aero Engine Supply Chain. The minimum requirements for both variable and attribute measurement systems are defined where used to validate product in its final condition.

It is expected that these requirements shall be flowed down to all sub tiers within the supply chain and included within their Quality Management System.

Any additions or exclusions to this standard must be agreed with the purchaser.

The scope of application of these requirements shall be defined by the purchaser, and this may include part numbers, processes, etc.

4. OVERVIEW OF MEASUREMENT SYSTEMS ANALYSIS (MSA)

A Measurement System is the combination of people, equipment, materials, methods, environment, analysis and decisions made on the measured results. All measurement systems have a level of uncertainty associated with them because of variation in these factors. MSA is the method of identifying the level of uncertainty of the whole system so that we can determine if the Measurement system is 'fit for its intended purpose,' i.e., the level of measurement variation is not significant (Table 2 shows the acceptance limits for MSA).

There are several types of MSA; which type is required will depend on the type of data being measured and the influences on the system (see Figure 2).

The purpose of MSA is to identify the total variation present in the system so that actions can be taken to effectively control it and ensure repeatable and accurate measurements. These studies should be conducted to represent the 'real world' as much as possible; e.g., range of inspectors, parts that cover the whole specification, normal working environment, etc.

MSA shall be conducted as part of New Product Introduction to validate the measurement system prior to production. There are also situations where MSA should be repeated, these include: changes to gauge design, refurbishment/repair, environment, product design change to the feature being measured, etc.

Table 1 describes situations where MSA is required. Note that in this context MSA refers to the methods described in Table 2. Which tests are applicable will depend on the type of data being measured and the influences on the system (see Figure 2).

This standard defines the minimum acceptance limits for MSA and provides guidance for most situations but there may be situations where alternative methods are required. In such cases they must be approved by the purchaser.

Table 1 - MSA applications

Event	Event Description	Action
1	New inspection device or method introduction.	Perform MSA
2	New/Changed Production Process.	Evaluate current or Perform MSA
3	Any significant change to the current inspection device or method: i.e., equipment, operator, environment, location, sequence, calibration standard, Inspection house, CMM software or hardware change	Evaluate current or Perform MSA
4	Following a product escape related to (or suspected to be) from the Measurement System (nonconforming material left the facility).	Evaluate current or Perform MSA
5	Change in how an inspection device or method is used, or its application*. For example: 1. When changing from simple geometry to complex. Moving from simple linear dimensions with flat parallel surfaces to non-flat (non-parallel) surfaces with geometric constructions required. 2. When changing from similar to non-similar product characteristics. Moving from visual inspection of edge breaks with dimensional requirements to visual inspection of cosmetic appearance requirements.	Perform MSA
6	Product requirements are changed to be more restrictive or tightened.	Recalculate from base data or Perform MSA
7	As part of a First Article Inspection (FAI) following a lapse in use of more than 24 months.	Evaluate current MSA
8	Existing inspection device or method is being used to accept product and has not previously been evaluated per this standard as directed by purchaser.	Perform MSA where required
9	Product audit non-conformance or product investigation when suspected to be from the measurement system.	Evaluate current or Perform MSA
10	To verify a measurement system is adequate before SPC.	Perform MSA where required

* Different product with similar geometry and tolerances are typically not considered a change in application.

NOTES:

- Specifications within this table apply unless otherwise stated by the purchaser
- The term “evaluate” means a confirmation that the MSA study characteristics are still valid and additional measurement uncertainty has not been induced

5. ORGANIZATION AND COMPETENCE REQUIREMENTS

The correct training of MSA practitioners is key to the successful outcome of the process. Each supplier shall employ or have access to a practitioner who has appropriate experience and can demonstrate competence that includes all elements of this standard. It is expected that the practitioner will periodically confirm compliance to the standard. Individuals involved in the study must be suitably trained and competent in the measurement task. They must be representative of the measurement system users (see 8.1.3). The MSA study must be lead or facilitated by a person trained and competent in the methodology covered in this standard.

The supplier shall nominate a suitably qualified and experienced person from within their own organization as accountable for deployment of this standard and respective compliance.

6. QUALITY SYSTEM

- The supplier shall have a documented process within its own quality system which meets the requirements of this standard. The process shall be fully implemented and subject to an audit.

- The documented MSA process shall describe the training and competency requirements for practitioners of this process.

- Records of MSA studies identifying measurement systems as capable shall be maintained within the organization's quality system and subject to the same level of record retention as its FAI Records.

7. MEASUREMENT SYSTEMS ANALYSIS REQUIREMENTS

7.1 Pre-Requisites

The pre-requisites and generic requirements for any type of MSA study are:

- The measurement equipment must be calibrated and traceable to a relevant national or international standard.

- The measurement equipment must be maintained in good condition and checked for evidence of damage or wear, which may impact the measurement capability.

- Production parts must be used for studies, except for circumstances where the use of representative parts or artefacts is authorized (see 8.1.1 and 8.1.2). The parts must represent the full tolerance and it is beneficial to include parts just outside the LSL and USL.

- The parts should be as clean and burr free as would be seen by the production inspection method.

- Individuals involved in the study must be suitably trained and competent in the measurement task. They must be representative of the measurement system users (see 8.1.3).

- The measurement system analysis study must be lead or facilitated by a person trained and competent in the methodology covered in this standard.

- An environment representative of the production operation must be used for the MSA study.

- The method used for any study must replicate the conditions in the production process. Where alignment, fixturing, and clamping could influence the measured value, the component must be removed and reloaded between each measurement. Deviation from this requires purchaser authorization.

- During the study, the personnel performing the measurements must not have visibility of either their own or other study participant's previous results.

7.2 Considerations when Planning a MSA Study

MSA Studies require careful planning to ensure that the results are truly representative of the measurement system. The system shall be fully evaluated to identify what could affect the results so that anything likely to contribute to the variation is included in the study. An MSA is essentially an experiment to determine the degree and causes of variation within a measurement system. As such, careful use of Design of Experiments is recommended. See also 8.1.

Factors that need to be evaluated include:

- Environment – temperature, humidity, contamination, vibration, electromagnetic radiation, etc.
- Location – different buildings, sites, etc.
- Part variation that will affect the measured value (surface finish, flexibility, shape, size, etc.)
- People – shift patterns, times of the day, experience levels
- Process – fixtures, probes, accessories, etc.

A useful way of visually expressing the factors is to use a tree diagram example shown below.

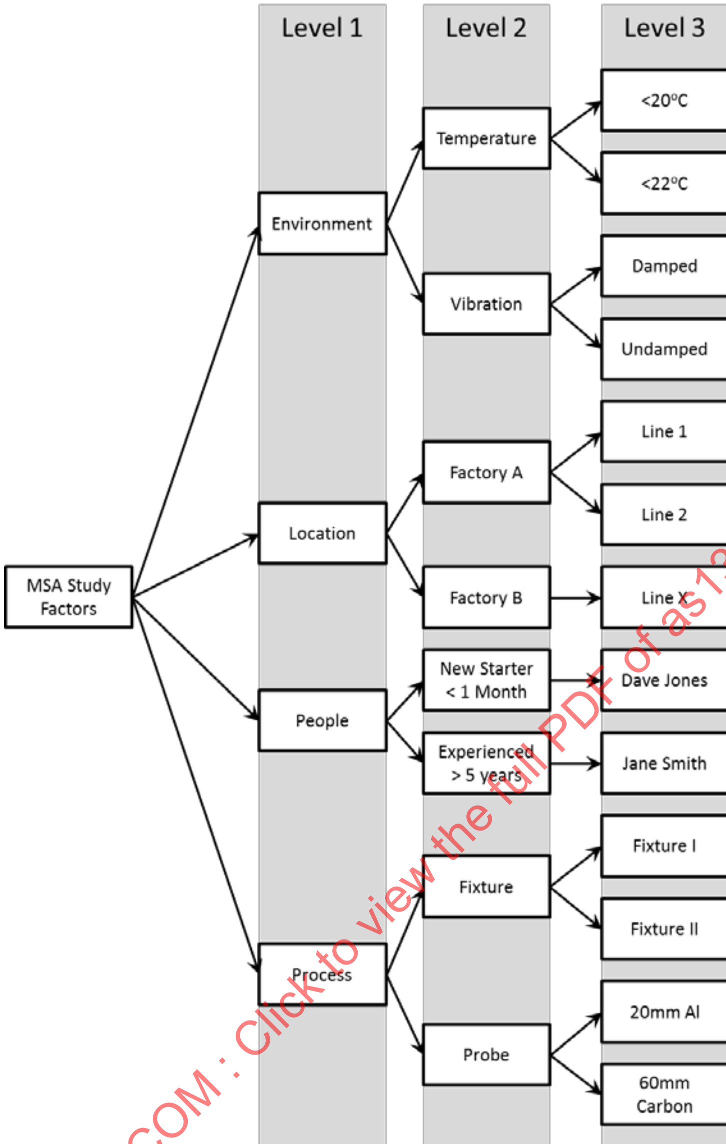


Figure 1 - Planning an MSA study

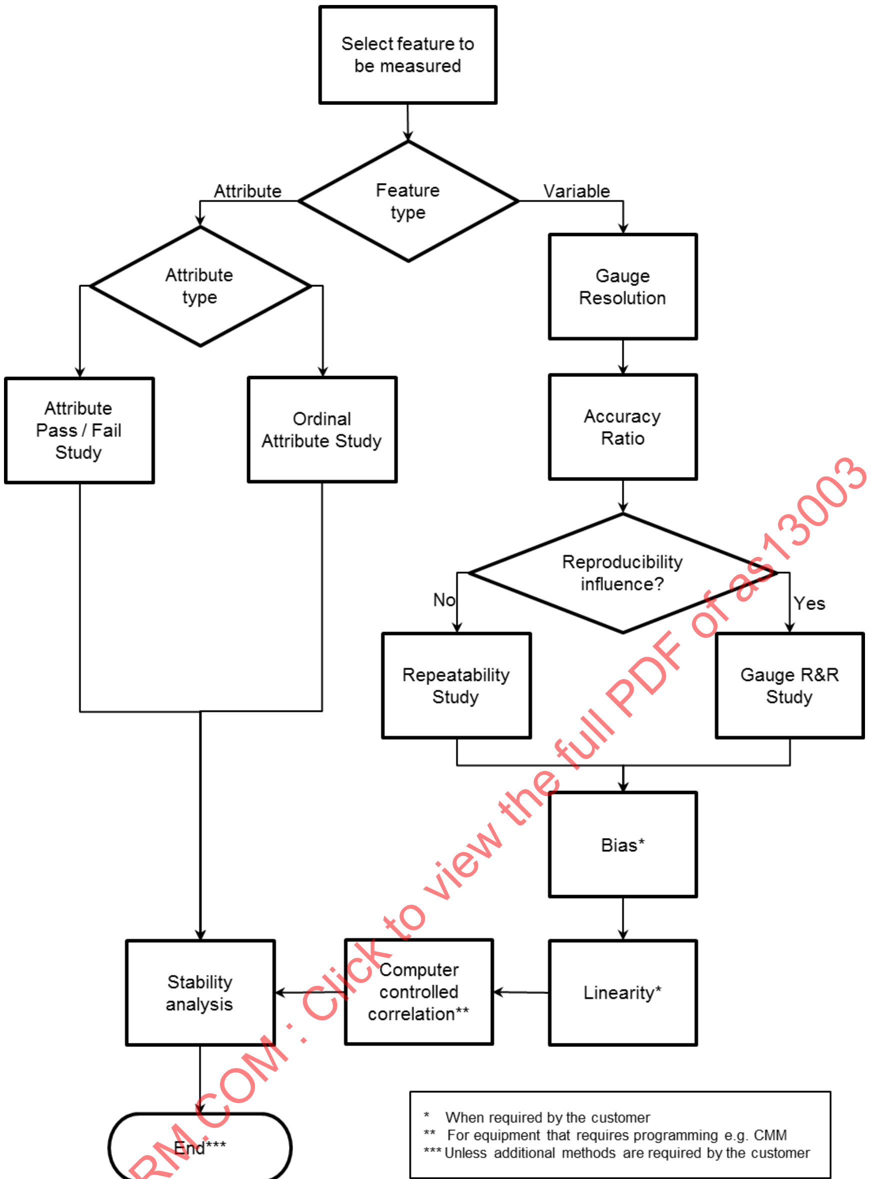


Figure 2 - Selecting the appropriate measurement systems analysis methods

Figure 2 shows methods of assessing MSA that can be applied in a logical order. The selection of methods should be based on purchaser requirements (i.e., not all tests may be required). The type of measurement being conducted and the study characteristics such as part availability, number of operators, number of repeat measurements, etc., many influence the selection of one method over another. The detailed descriptions and case studies within this standard can guide the user to the selection of MSA methods. See 7.2 of this standard.

7.3 MSA Minimum Requirements

Table 2 - Minimum requirements for MSA

Method	Feature Category			Comments
	Critical	Major	Minor	
Resolution	≤10% of total tolerance ***			Based on total tolerance.
Accuracy ratio**	Requirement = 10:1		Requirement = 4:1	Values up to 4:1 may be acceptable when approved by the purchaser
Accuracy Error / Bias	≤10% of total tolerance			Purchaser requirements may override this
Repeatability	≤10% of total tolerance	≤20% of total tolerance	≤30% of total tolerance*	Purchaser requirements may override this
Gauge R&R	≤10% of total tolerance	≤20% of total tolerance	≤30% of total tolerance*	Purchaser requirements may override this
Computer driven measurement systems correlation	≤10% of total Tolerance		≤20% of total Tolerance	Purchaser requirements may override this
Linearity**	≤1% of total tolerance		-	
Attribute Study: pass/fail	Kappa ≥ 0.8		-	Only required on operator dependent interpretation
Attribute study: ordinal	ICC ≥ 0.75		-	Only required on operator dependent interpretation

* Repeatability and Gauge R&R studies are only required for minor features when there has been an escape related to that feature or when required by the purchaser (see Table 1).

** Only required when specified by the purchaser or where the feature is a datum.

*** May permit lower resolutions provided an acceptable GR&R is obtained.

NOTE: All statistical calculations assume ±3 standard deviations unless otherwise specified by the purchaser.

In cases where high process capability are available, best practice is to check measurement capability against the “% of process variation” rather than the “% of total tolerance” seen in Table 2. As process variation is less than the tolerance, applying these limits will ensure the measurement system adequately controls the process and the manufacturing process is not limited by the measurement system.

8. ELEMENTS TO CONSIDER WHEN CONDUCTING MSA

8.1 MSA Design Factors

The following section are designed to provide guidance when conducting MSA studies. Where this guidance cannot be followed as written then alternatives may be appropriate and in such cases this should be agreed with the purchaser.

MSA studies should be designed to find all of the existing variation in the measurement process so that action can be taken to mitigate it and provide a capable and repeatable measurement system.

8.1.1 Sample Selection

Factors affecting the selection of samples include:

- Criticality of dimensions – the degree of confidence required for critical dimensions is higher, therefore, more data is required.

- Part configuration – large parts, inaccessible features and low numbers of available samples may dictate low sample sizes which must be recognized in any reports.

- Purchaser requirements – specific requests regarding the selection of samples may be defined by purchasers.

Sample parts shall represent the entire production operating range and ideally the entire allowed tolerance range.

The analysis techniques used assumes that there is statistical independence of the individual data points so must be taken to randomize the measurements ensuring that operators are not able to identify parts.

In some instances it will not be possible to obtain a fully representative sample of product. In these instances, a feature that is representative of the manufacturing process and size can be used. Where a representative part or linearity study is used, to compensate for a poor sample of parts, this must be documented and approved by the purchaser.

8.1.2 Sample Numbers

As with any statistical technique, the larger the sample size, the more accurate the results. The number of samples is, however, very dependent on the measurement process and the type of study being conducted. Studies where the measured characteristic has significant variation will require a high number of samples to statistically describe the process, 10 samples or more are required. Where this cannot be achieved, it should be declared as part of the MSA acceptance report.

Processes where the measurement system is subject to human intervention will require a higher number of repeat measurements per inspector. Each inspector shall measure each sample three (3) times or more. As a minimum and where samples show a good level of control, each person shall measure each sample twice.

Attribute studies require a much larger sample size, typically 30 or more.

Where automated measurement systems are used (CMMs, etc.) the human influence should have negligible effect. The number of repeats can, therefore, be reduced to a minimum of five (5) when proving repeatability of automated measurement systems or where the operator has no influence on the measured results (see case study 10.1 in this standard).

8.1.3 Operators

Where possible, a representative sample of operators who normally use the measurement equipment shall be included in the study. If this is not possible then a minimum of two (2) people shall be selected. Important factors to consider include:

- The most and the least experienced people
- People who work in different locations
- People who work on different shifts
- Any physiological factors that may affect the measurement process; e.g., left and right handed people, different levels of eyesight acuity, height, strength, etc.

All participants in the measurement study shall be trained and experienced in the task. Do not include people who do not normally carry out the measurement activities as part of their day-to-day duties. Do not include 'experts' like Metrologists or Manufacturing Engineers unless they are representative of typical users.

8.1.4 Low Sample Sizes

Ideally at least ten parts or features should be available for the measurement study. Using three operators measuring each part a minimum of two (2) times would give sixty data points, which is sufficient to give meaningful result.

Sometimes though there may not be sufficient parts available. There are a number of ways that this problem can be overcome:

- Conduct the study over a longer period of time as parts become available. This also has the advantage of incorporating any time based factors into the study.
- Use multiple features on one (1) part. For example, a slotted disc may contain several identical features which can be identified and used to conduct the study. Potential problems with using this method include randomizing the measurements effectively and a lack of variation in the features, but with care this can be accomplished. Be aware that this method may exclude sources of variation that could be affecting the measurement values and any results obtained may not be fully representative.
- Use scrap parts. So long as the part is representative of a finished part (all features are present and the geometry and surface finish are correct) then non-conforming parts can be used. Ideally use features on the parts that are conforming in case that non-conformance affects the measurement.
- If no other options are available then smaller sample sizes may be used but a note should be added to the final report highlighting the small sample size and higher risk of the results being unrepresentative. Care should be taken as a low sample can also lead to a reported low number of distinct categories which may or may not be a true problem.

8.1.5 In Part Variation

When conducting a measurement study, it is important to identify variation in the part that could affect the measurement result. As an example consider measuring a shaft diameter where the diameter of the shaft is not perfectly round. Running a measurement study will capture both variation due to the measurement system, and variation of the diameter shape. It is advisable to identify the amount of in part variation before the study is conducted. Marking the component to ensure measurements are taken in the same place will limit the in part variation, but the study will not then be representative of the measurement process seen in production. Where in part variation is considerable, compared to the size of the tolerance, this should be declared to the purchaser as part of the measurement study.

8.2 Analysis of Results

After choosing the most appropriate sampling plan for collecting the measurement values, an appropriate analysis method shall be selected to interpret the results.

In order to progress with statistical studies the following steps describe some quality checks to ensure the data has been correctly gathered:

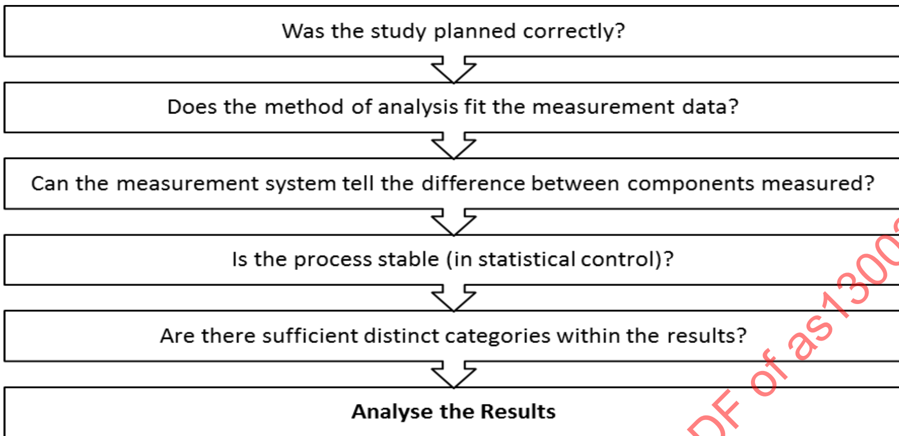


Figure 3 - Analysis of results flowchart

8.2.1 Gauge Resolution

The resolution of a gauge is an inherent property of that instrument and is usually fixed by its design. It is the smallest readable unit or usable output from that instrument. Care should be taken not to assume that the smallest increment on the gauge indicates its resolution, as this is often not the case. The gauge manufacturer should specify the least-significant digit (LSD) and it is this that should be used to test the gauge's suitability for its application.

Resolution may also be affected by 'noise' such as electromagnetic interference, vibration, friction, etc., as well as the physiological limitations of people reading the gauge; e.g., eyesight.

For example: a depth gauge is used to measure the length of a location hole on a casting. The hole size varies by 0.1 inch, which is also the drawing tolerance allowed for the feature. The gauge can be read to 0.01 inch by an operator with average eyesight. This means that the resolution is 0.01 / 0.1 * 100 = 10% which is acceptable.

8.2.2 Accuracy Ratio

This is calculated by dividing the total part feature tolerance (from the part drawing or specification) by the total calibration tolerance spread of the measurement equipment. For example, if the tolerance for the diameter of a shaft is quoted on its drawing as 15.6 mm ± 0.05 mm then the total tolerance is 0.1 mm. This is measured by a micrometer which has a calibration tolerance of 0.003 mm. This gives an accuracy ratio of 33:1 which is acceptable.

8.2.3 Repeatability

A repeatability test, also known as a Type 1 Gauge Study, will identify the variation observed when one operator performs repeated measurements on one part with the same instrument. The best way to analyze the results of this type of study is to plot the values on a graph.

In this example (see Figure 4) a feature has been measured 50 times and the results plotted on a run chart.

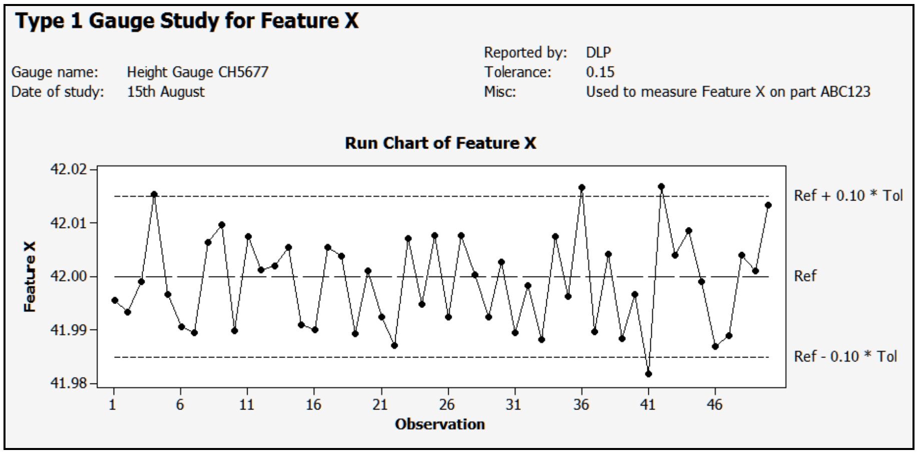


Figure 4 - Type 1 gauge study

The accepted reference value of 42.0 mm has been added as a reference line (‘Ref’ on chart) so that we can check for bias.

Lines have also been added to the chart to indicate 10% of the tolerance (‘Ref +/- 0.10*Tol’ on chart). Ideally all results would be within the 10% limits, but in this example four (4) are not so the source of this variation should be established and eliminated before the measurement system can be accepted.

8.2.4 Gauge Repeatability and Reproducibility (GR&R)

This is the most common MSA study used and determines how the variation in measurement system is split between repeatability and reproducibility. In a good measurement system, the largest variation obtained is due to part to part differences not variation due to the measurement system.

Once the study pre-requisites (7.1) and planning (7.2) are completed then the study can be run and the data collected and analyzed, ideally using a suitable statistical analysis software package such as Minitab (contact your purchaser for details of approved software packages).

There are two approved methods for analyzing the data: an ANOVA (analysis of variance) method and an Xbar and R method. The calculations used in Xbar and R method are simpler but the ANOVA method is recommended.

The objective in the analysis of a Gauge R&R study is to split the variation into individual components: repeatability, reproducibility, parts, operators, etc., and then look for indications of any specific sources of variation. The results are easier to interpret if they are expressed graphically (see Figure 5).

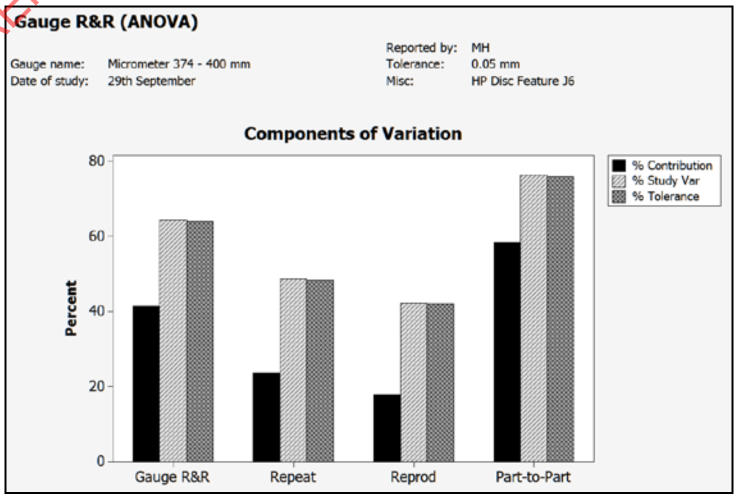


Figure 5 - Components of variation

The sources of variation are split into component parts and are expressed as percentages; total gauge R&R (left hand sub-group), gauge variation split into repeatability and reproducibility (center two) and also part-to-part, which is the variation present in the parts being measured . Each sub-group is expressed in three different ways (represented by the different shading in the example); as % contribution and then individual components divided by the total study variation and finally by the tolerance defined for the part under test (see Table 2 for minimum requirements).

If any individual elements are greater that the requirements then further analysis will be required to try and establish why.

Another useful way of expressing the data is to use an Xbar and Range chart which plots the mean average values of each part measurement by the range of values for that particular subgroup. If operator A measures part 1 twice, the average of the two values along with the difference between the two measurements are plotted (see Figure 6).

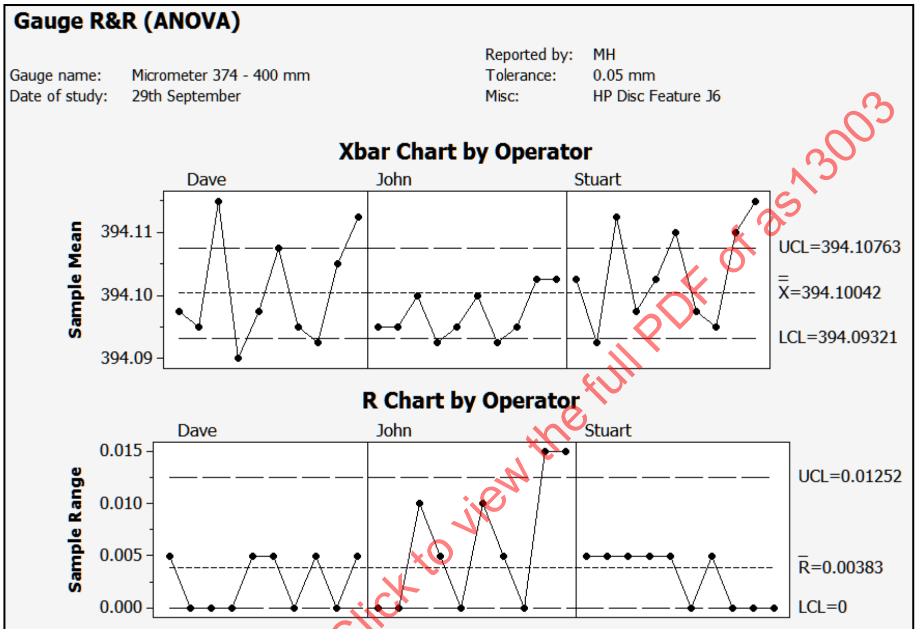


Figure 6 - Xbar & R charts

Because the parts chosen should represent the entire range of possible parts the graph should ideally show a lack of control; i.e., points outside the horizontal dashed control limit lines (UCL and LCL). In the Xbar chart the lack of control actually shows that there is more part variation than measurement variation.

This chart also allows a comparison of the different operators and can show how good the discrimination of the gauge is. In this example we see that operator John is producing different results from the other two and has roughly three times the variation in his repeat measurements (up to 0.015 mm compared to 0.005 mm). It can also be seen in the example John's 9th and 10th measurement points are out of the control limits in the R Chart by Operator. This gives cause for concern and the study may be rejected due to the lack of control. Without stability in the range of measured values the measurement error is not predictable and could lead to a false study result. Instances where a lack of control is detected can be recorded as part of the measurement system study or through the inspection limitations process and reported to the purchaser, however it is better to review the process and John's training to ensure control of the process is established before repeating the trial. The discrimination of the gauge can also be seen to be 0.005 mm by observing the steps in the data plots (0.000, 0.005, 0.010, etc.).

A number of other graphs can be produced to illustrate the data in useful ways. If reproducibility is significant then the data from each of the operators can be plotted to show the average measurements for each part. Any patterns in the data can then be identified to help determine the cause of the variation.

In this example (Figure 7), operator John measures seven (7) out of the ten (10) parts smaller than the other two individuals.

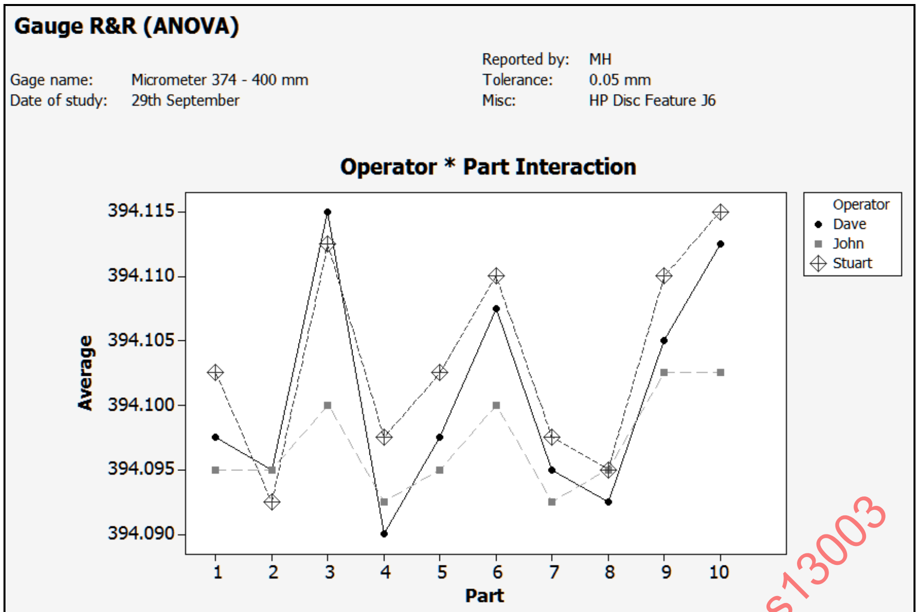


Figure 7 - Operator * part interaction

He also seems to have a problem with measuring dimensions larger than 394.105 mm where his variation is much larger.

For measurement system accept/reject purposes the figure that should be used to compare against the values in Table 2 is the % Tolerance value, i.e., the total variation percent of tolerance for each component.

Source	***Standard Deviation (SD)***	***Study Variation (6 x SD)***	***% Study Variation (% SV)***	***% Tolerance (SV / Tolerance)***
Total gauge R&R	0.0053424	0.0320546	64.49	64.11
Repeatability	0.0040311	0.0241868	48.66	48.37
Reproducibility	0.0035059	0.0210357	42.32	42.07
Operator	0.003092	0.0185517	37.32	37.1
Operator*Part	0.0016527	0.0099163	19.95	19.83
Part-To-Part	0.0063319	0.0379912	76.43	75.98
Total variation	0.0082846	0.0497075	100	99.41

If the tolerance (Upper spec - Lower spec) is given, percent tolerance is calculated by dividing the Study Variation for each component by the specified tolerance.

The Study Variation is the standard deviation component calculated in the study multiplied by six (6) which is the number of standard deviations needed to capture 99.73% of the process measurements.

8.2.5 Linearity

Linearity should be evaluated through the entire range of dimensions that the measurement system for which it is likely to be used. The minimum number of points required to do this is three (3): minimum and maximum dimension plus a mid-point to check for consistency of any error. Ideally 5 will be used in order to give more reliable results. For example (see Figure 8), if the range of measurements is 25 to 45, then measurements should be made at 25, 30, 35, 40, and 45 °C.

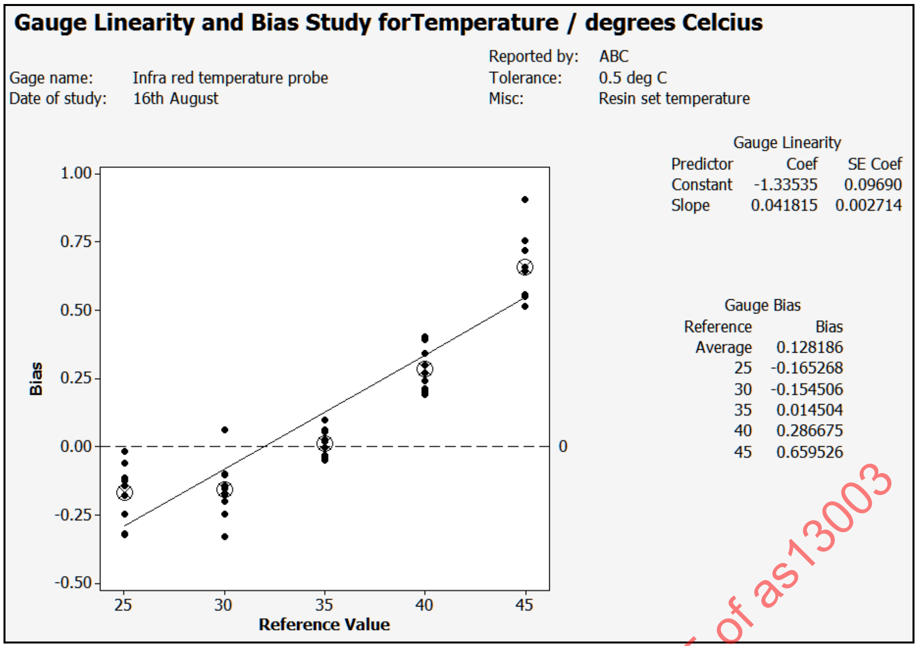


Figure 8 - Linearity and bias

The key to evaluating linearity is to be confident that the true values at each of the measurement points are known, which is not always easy to do. Generally, using an alternative proven method to establish true values is the only way to do this.

Each point should be measured at least 10 times by one operator. Ideally do this in random order so that the operator does not influence the result. Using five (5) points this would give a total of 50 measurements.

The bias observed at each of the measurement points can then be calculated by subtracting the reference value from the recorded value and then plotted on a graph (the black dots in the example in Figure 8). The average bias at each point may then be calculated (the circled X in the example). If there was no bias the average values would all be on the zero dashed line.

The best fit line can be determined by (the solid line on the chart) using a software program that does this for you.

In this example, the bias at reference value = 35 °C is acceptable but below this the temperature gauge under-reads. As the reference temperature value increases the bias also increases up to a maximum of 0.659 over-read at reference value = 45 °C. This demonstrates a lack of linearity and shows that full gauge R&R studies must be performed at all reference values. The slope of the best fit line may be calculated so that bias levels may be predicted at other values. Do not, however, extrapolate this line outside the range chosen as the error may not be consistent.

8.2.6 Stability

Stability, sometimes referred to as drift, is a measure of the measurement system variation over time. This is sometimes caused in instruments by deterioration in mechanical or electronic components, or by a change in the method used by the operator, for example by trying to do the task more quickly.

This is best measured by using a process control (SPC) chart (see Figure 9).

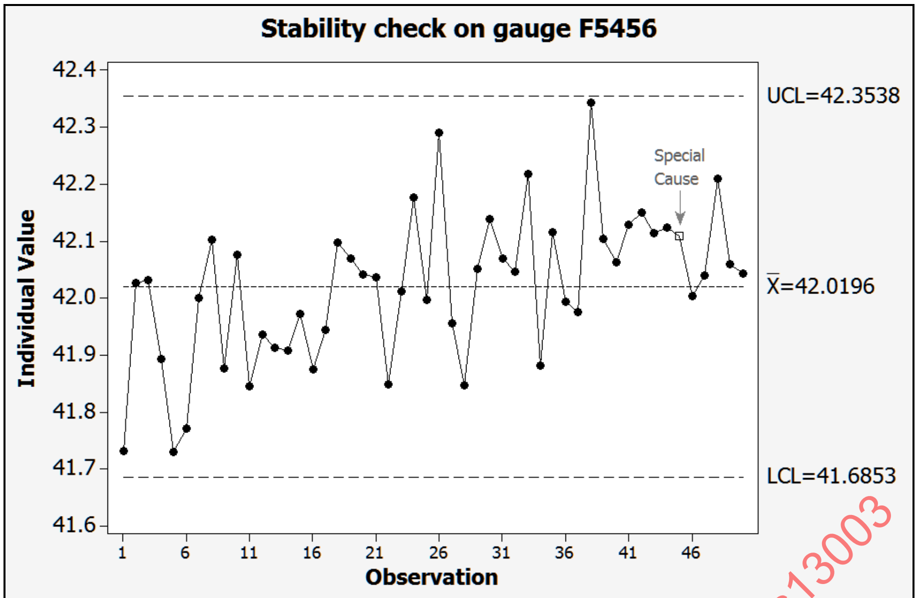


Figure 9 - Process control chart

To carry out a stability check a reference sample needs to be measured repeatedly over a period of time. Some variation will normally be observed due to the standard measurement error associated with the system but there should not be any trend in the data or evidence of any special causes of variation (indicated on the chart in Figure 9). Any sign of instability should be investigated and the cause identified and removed. A stability check is often used for Artefact testing on measurement systems such as CMMs. Tests the measured result of a known item every shift/day/week to confirm the measurement system continues in a known status between calibrations. The use of Process Control Charts is explained in greater detail in the reference documents detailed in Section 2.

8.2.7 Measurement Systems Bias

Measuring a feature with two (2) different measurement systems will sometimes give two (2) different measurement results. The relationship between the two (2) values is the measurement system correlation. To ensure measurement systems give accurate answers, measurement correlation is often checked against a reference value which can be considered as a very accurate measurement. This is often known as the measurement bias.

Where a number of measurements are taken (best practice is to take ten or more), the average value of the range of measurements is used to evaluate the bias between the measurement and the reference value. Bias is normally quoted as a direct value (in the same units as measured), as a percentage of the overall process variation or as a percentage of the allowed tolerance. An unacceptable level of correlation or bias would indicate that the measurement system has an error, the measurement process is flawed or there is a calibration error.

NOTE: Bias is often evaluated at different dimensions in the range of the measurement system; see section on Linearity in 8.2.5.

8.2.8 Measurement System Discrimination (Number of Distinct Categories)

The number of groups within the process data that the measurement system can discern is used as a quality check of the measurement system; if the number of categories is low, the measurement system might be poor, or the measurement samples are clustered compared to a relatively large tolerance zone. If a measurement system's discrimination is inadequate, it may not be possible to accurately measure process variation or quantify measurements for individual parts.

Table 3 - Number of distinct categories

Number of Categories	***Can Provide***
1	Poor information limited to conformance versus non-conformance.
2	The data can be divided into two groups by the measurement system and is therefore insensitive for control of a process as only high and low values can be identified.
3	The data can be divided into 3 groups representing high values, low values and mid values. This is of limited use for detailed process control.
5 or more	Demonstrates the ability to provide good process control, as measurement data can be split into one of 5 groups through the data range.

A value below five (5) may indicate that the parts measured as part of the study are too similar or do not represent the entire range of the process. A low number of categories may also indicate that the measurement system precision is poor and needs improving.

8.2.9 Attribute Data

Where attributes are used to assess feature acceptability (e.g., do people make the same decisions of pass/fail criteria) the evaluation of measured variable data is not possible. Typically there are two methods used to assess attribute data:

<u>Attribute Agreement Analysis or Pass/Fail Study</u>

This is a test of three (3) criteria:

- Do assessors make the correct decision when evaluating a feature's acceptability?
- Are they consistent in making that decision? (i.e., if they did the evaluation again would they give the same answer?)
- If there are multiple assessors, do they all make the same decision?

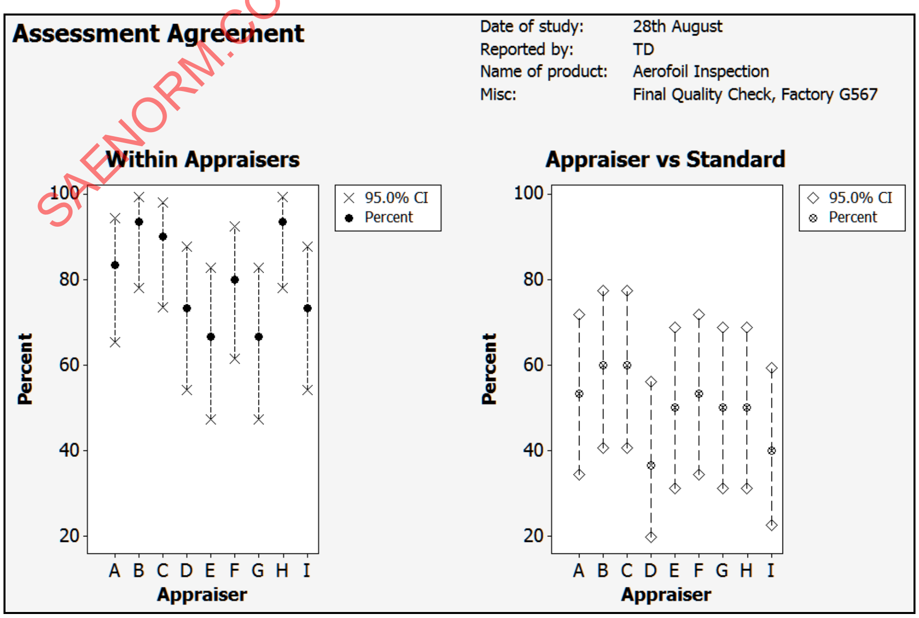


Figure 10 - Attribute agreement analysis

Results can be expressed as the number of correct/incorrect answers and also as percentage% values. In this example (see Figure 10) we can see that two appraisers (B and H) have a 'within appraiser' or consistency score of 93% but two others only have a 67% score (E and G). Against the standard, the best result is 60% which means that 40% of the time they give the wrong answer. It is also possible to calculate a value called Fleiss' kappa statistic, which is a measure for assessing the reliability of agreement between a fixed number of assessors. If kappa = 1, then there is perfect agreement. If kappa = 0, the agreement is the same as would be expected by chance. The stronger the agreement, the higher the value of kappa. Negative values occur when agreement is weaker than expected by chance, but this rarely happens. Depending on the application, a kappa value less than 0.7 indicates that the measurement system needs improvement against the target value (Table 2) of 0.8. Kappa values greater than 0.9 are considered excellent. In the example in Figure 10, the best kappa value that was achieved was 0.36 so there are major problems here that need to be addressed.

NOTE: Consult the guidance material in Section 2 for further details.

Attribute Study – Ordinal Data

The Interclass Correlation Coefficient (ICC) compares several different scenarios and uses sums of squares to accomplish this task. The main issue with an ICC is determining the reliability of ratings if the ratings are from a single assessor or if the ratings are averaged across several people. The interpretation of ICC is equivalent for all basic forms (each appropriate for a different situation). In this case the criteria acceptance ICC must be higher to 0.75 to obtain an almost perfect agreement between assessors.

8.3 Mitigation Strategies for Failed Results

Where the measurement/inspection system has failed to reach acceptable minimum levels (as described in Table 2), the measurement system may still be judged as fit-for-purpose if the relevant purchaser technical authorities agree. Specific situations where this may occur include:

- Significant capital investment is required to achieve the minimum standard level of capability such as the purchase of new equipment, improved environmental conditions or technology development.
- The size and tolerance of the feature determines that it is not possible to achieve the minimum standard level of capability. This is typically caused by very tight tolerances or a very large component that will not fit on normal measuring equipment.
- MSA is limited by component characteristics such as a flexible part, size or weight, thermal growth, etc.
- The constraints on the measurement study do not allow for a robust result. This may include a lack of suitable components, a lack of operators or a change in measurement environment, etc.

8.3.1 Limitations of Measurement

In order to judge if the measurement system is acceptable, the details of the measurement process, the measurement system analysis study details and the results achieved shall be submitted to the purchaser for verification and acceptance/rejection of the level of capability demonstrated. An example form is included in Appendix B. This can be included at FAIR stage for purchaser approval or in line with other purchaser requirements.

NOTES:

1. It is expected that the supplier will make every reasonable effort to prove the measurement system capability before submission of any measurement limitations to the purchaser.
2. The supplier can use historical measurement capability results to demonstrate that a purchaser's specification cannot be achieved and negotiate an improved tolerance, or increased acceptance limits from Table 2. This will be on a feature-by-feature basis.

8.3.2 Mitigation for Poor Measurement

To mitigate the effects of poor repeatability it is possible to take a specific measurement a number of times and then calculate the mean in order to estimate the true value. In this example (see Figure 11), 20 measurements have been taken of a single feature.

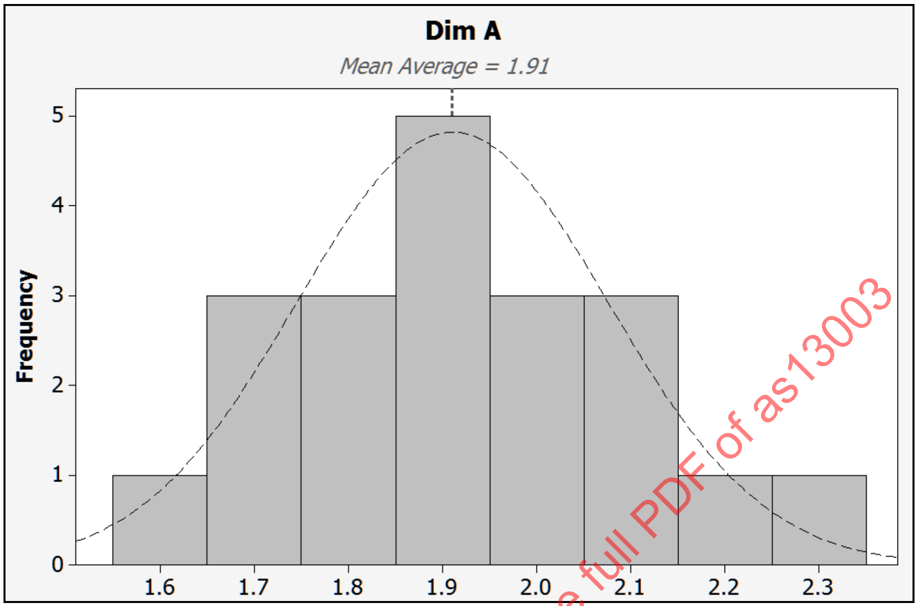


Figure 11 - Histogram of repeat measurement results

The 20 measurements range from 1.6 to 2.3, with a mean average of 1.91. Any measured value between 1.6 and 2.3 could be achieved from a single measurement, but statistically the true measured value for the feature is likely to be around the average reading and this could then be used to sentence the component. The number of times that a measurement needs to be taken in order to have confidence in the average value depends on the amount of variation present and the required accuracy – see the guidance referenced in Section 2 for more details.

Where poor measurement reproducibility is seen in a measurement study, it is likely that an element of the measurement system has produced different or variable measurement results. This can be operator, cell, instrument, or time related. Where an operator related issue is suspected, improved training or dedicated instructions can be implemented to remove the variation. It is important that the source of variation and the reproducibility study results can be used to track the root cause.

NOTE: Acceptances of measurement system mitigations such as averaging shall be documented and approved by the purchaser prior to measurement system use.

8.3.3 Fixtures and Flexible Components

Where fixtures are used to secure flexible components for measurement, the fixtures must be part of the measurement capability study. If there are multiple fixtures, variation in the fixtures may cause measurement variation and it is expected that this will be evaluated as part of the study.

Where fixtures affect the form or size of the component for measurement, it is expected that the fixtures will be under calibration control. This will ensure that any wear, damage or movement of the fixtures is maintained within acceptable limits.

If the component features are flexible, the measurement result will have variation due to both measurement process and part movement. It is important to recognize both sources of measurement variation and ensure the component is correctly sentenced. For the purpose of measurement with fixtures, trials can be limited to study repeatability and then establish measurement system bias.

As a simple test, measure the component at least three (3) times with the minimum of part movement or variation in the process to establish a base line measurement repeatability figure (the range of measured results between the lowest measured result and the highest measured result).

Remove the component from the fixture and measurement system and reset the component to mimic a new component set-up. Re-measure the component and repeat the process a further three (3) times minimum.

- The first group of runs, where the component was not removed from the fixture will estimate the measurement system repeatability.
- The second group of runs where the part was reset in the fixture will estimate the repeatability through part movement but will also include the measurement repeatability.
- Studying all the runs together will estimate the reproducibility of the full measurement process.

NOTE: This is not a statistical solution and is only used to indicate the source of measurement variation. Gauge Repeatability and Reproducibility and bias studies are recommended to prove measurement capability with flexible components and fixtures.

Where fixtures are used, the fixture may induce a measurement bias. Experiments shall be conducted to ensure the component characteristics are not unduly effected. This may include the measurement of the component using a number of different measuring systems, both on and off the fixture to establish the range of measurement obtained.

In all cases, component constraint must follow the purchasers' requirements or be documented and agreed through the measurement limitations process.

8.4 Environmental Factors

The environment in which a measurement system is used will impact measurement capability. Examples of measurement variation due to environmental factors include:

- Thermal expansion of the component
- Thermal expansion of the measurement system
- Expansion of the part due to humidity
- Assessment of color or surface finish in varying light conditions (florescent lighting, tungsten lighting, day light, etc.)
- Dust and dirt or oil contamination of measurement equipment or component surface
- Air movement and dust contamination when taking measurements with lasers

While the MSA study will detect some of the variation due to environmental conditions, the study will not take place over a long enough time period to fully account for all environmental impacts. In order to accurately determine environmental effects, the following steps are recommended:

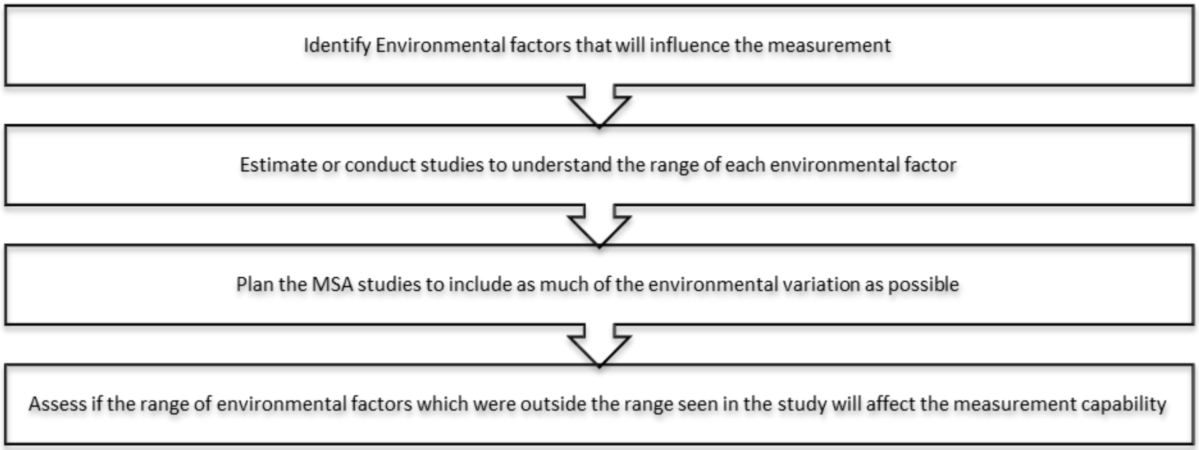


Figure 12 - Assessment of environmental impacts

1. The environmental conditions that will affect the measurement process shall be identified.

2. Each environmental condition needs to be assessed and/or measured. For example, a temperature study of the machine shop over a full year to understand the temperature variation that is seen in winter or summer periods where measurement is conducted.

3. Including environmental variation in the MSA study will ensure further compensation is not required

4. The effects on the measurement result shall be calculated and assessed. For example: an MSA study on dimensional measurement had a temperature variation of between 18 to 22 °C, however, the temperature variation where the measurement is taken varies between 16 °C and 26 °C over the full year. The MSA study results have been compromised by the increased range of temperature seen over the full year. (MSA study temperature variation = 4 °C: Machine shop yearly temperature variation = 10 °C). It is expected that variations of this nature that have a significant effect on the measurement process will be taken into account. This may be through the estimation of thermal growth of the component/equipment or the declaration on the deviation in the measurement study documentation or limitations process.

In all cases, MSA acceptance limits set within this standard must not be exceeded when the calculated effects due to environmental variation are added to the measurement capability achieved in the MSA study.

8.5 Use of Read Across Results in MSA

MSA studies are usually conducted on a specific feature, component and measurement system. In certain circumstances it may be appropriate to read across measurement capability results in place of conducting a new study. This may only be done when the measurement system characteristics are judged to be a suitable equivalent. The decision to read across capability shall be confirmed by assessment of study characteristics from the donor study to the system under test, taking into account feature criticality.

Acceptance of read across must be documented and approved by the purchaser authorities through the inspection limitations process. Examples of study characteristics and acceptance criteria are indicated in the table below:

Table 4 - Applicability for MSA read across

Characteristic of Control and Definitions	Critical Features	Major Features	Minor Features
Type of equipment *Is the same type of measurement equipment used? (Digital, Vernier scale, gauge construction, etc.)*	Equipment must be identical		Equipment must not induce additional error
Measurement resolution *Is the minimum scale available to read on the measurement device the same?*	Must be the same or better		Must be the same or better
Feature tolerance *Is the bandwidth between the upper and lower feature acceptance limits the same?*	Recalculate MSA on new tolerance		
Feature Type *Is the manufacturing process or shape of the part the same (circular, prismatic, free form, etc.?)*	Identical process and characteristic		Similar characteristic
Feature form, surface finish, flexibility *Does the feature have the same manufacturing characteristics? (Part flexibility, robustness, surface finish, surface lay, etc.)*	Same surface finish, form and feature flexibility	Similar surface finish, Similar form, the same or better flexibility	Feature flexibility must be similar
Measurement environment *Are the environmental effects around the measurement system that may change the size of the gauge or component the same?*	Must have the same environment control	Environment monitoring indicates low risk to measurement	n/a
Operator competence *Does the Operator influence the measured result?*	Must use operators with similar competence		
Feature accessibility *Can the measurement be easily taken or is access to measurement difficult?*	Donor and target will have the same access		n/a

9. AUDIT CHECKLIST FOR MSA

See Appendix A.

10. MSA CASE STUDIES

10.1 Case Study - Resolution

The drawing requirement for a compressor case diameter is 10.312 inches ± 0.002 inches. This is a critical characteristic. What is the required gauge resolution for the measurement equipment?

The measuring equipment must be able to discriminate to at least one tenth of the total tolerance being measured, per Table 2.

In this case, for the total part tolerance range of 0.004, to meet the 10: 1 resolution requirement, 0.004/10 = 0.0004. The gauge must be able to measure with a resolution of 0.0004.

Trailing zeros when calculating limit dimensions from plus or minus dimensions may be disregarded provided all other requirements of this instruction are met. Example 0.0425 ± 0.0075, may be treated as 0.035 - 0.050 (3 place decimal) versus 0.0350 - 0.0500 (4 place decimal).

10.2 Case Study – Accuracy Ratio

The drawing requirement for a combustion case diameter is 16.714 inches ± 0.002 inches. This is a critical characteristic. What is the total measurement equipment accuracy required to maintain proper accuracy ratio?

The required accuracy ratio may be obtained by taking the total tolerance spread of the characteristic to be measured and dividing by ten (10), per Table 2. This will provide the largest acceptable total gauge accuracy permitted for that particular dimension.

In this case 0.004/10 = 0.0004. To maintain the 10:1 accuracy ratio, the total calibration tolerance spread of the Measurement Equipment must be 0.0004 inches or less.

10.3 Case Study – Repeatability

As part of the validation for the measurement of control blades, a mini repeatability study was conducted to validate two dimensional features:

A coordinate measuring machine (CMM) equipped with a 5-axis scanning head (continuous probe movement over the part surface to gather many data points) was used. The CMM was programmed in line with the part definition and according to the purchaser methodology that defines the method of programming and the datum points to use. The component is loaded in a fixture to locate the part for inspection. The datum system is made directly on the blade surface so several iterative calculations are required to generate a repeatable and accurate CMM datum system (CMM measured points need to be taken on the datum points specified on the drawing, so the datum is measured and recalculated several times to ensure the correct surface location is measured). The fixture is therefore not a factor in the measurement system.

For this study, a sample of three (3) parts were selected that represent the expected range of the process variation (a part on maximum metal condition, a part on minimum metal condition and a part around mid-limit), each part was measured 10 times by the same operator on the same fixture. Note: The changes in measurement setup are limited to ensure variation from other factors is not introduced between runs. The range method was used for calculation based on the variation of the three runs by part and analyzing the worst case of the three (3) parts to provide a quick approximation of measurement variability.

In this example, only two geometrical dimensions are studied but the same process would be used for the entire geometrical definition:

Part n°1	Dimension n°1 17.2 ±0.025	Dimension n°2 58.15±0.02	Part n°2	Dimension n°1 17.2 ±0.025	Dimension n°2 58.15±0.02	Part n°3	Dimension n°1 17.2 ±0.025	Dimension n°2 58.15±0.02
Run1	17,221	58,166	Run1	17,222	58,166	Run1	17,229	58,156
Run2	17,228	58,165	Run2	17,228	58,168	Run2	17,228	58,154
Run3	17,223	58,168	Run3	17,230	58.168	Run3	17,232	58,158
Run4	17,225	58,167	Run4	17,223	58.167	Run4	17,230	58,158
Run5	17,227	58,168	Run5	17,224	58.166	Run5	17,231	58,157
Run6	17,224	58,165	Run6	17,229	58.168	Run6	17,229	58,156
Run7	17,222	58,166	Run7	17,226	58.167	Run7	17,228	58,154
Run8	17,223	58,167	Run8	17,223	58.167	Run8	17,232	58,158
Run9	17,223	58,168	Run9	17,222	58.168	Run9	17,230	58,157
Run10	17,226	58,166	Run10	17,228	58.166	Run10	17,228	58,157
Max Range	0,007	0,003	Max Range	0,008	0,002	Max Range	0,004	0,004
IT	0,050	0,040	IT	0,05	0,04	IT	0,05	0,04

CALCULATION RANGE METHOD:

$$\%\ Repeatability = 100 * \left(\frac{Max\ Range}{IT}\right)$$

The calculation is conducted on all the geometrical dimensions for all the runs and this can be easily completed in spread sheet software package.

RESULTS:

Part n°1	Dimension n°1 17.2±0.025	Dimension n°2 58.15±0.02	Part n°2	Dimension n°1 17.2 ±0.025	Dimension n°2 58.15±0.02	Part n°3	Dimension n°1 17.2 ±0.025	Dimension n°2 58.15±0.02
% Repeatability	14%	8%	% Repeatability	16%	5%	% Repeatability	8%	10%

ANALYSIS OF RESULTS:

The maximum repeatability observed across all parts is used to assess if the repeatability is acceptable by comparison with the minimum acceptance criteria set out in Table 2 of this standard. The result for this case study can be presented using the table below:

	Maximum % Repeatability	Minimum Requirements	Acceptance
Dimension n°1 17.2 ±0.025	16%	< 30 % Feature Minor	OK
Dimension n°2 58.15±0.02	10%	< 20 % Feature Major	OK

It is observed that the repeatability for dimension N°1 part 2 gives the maximum value of 16% of the total tolerance. As this is below the minimum standard established for "Minor" feature category (Repeatability <30% from Table 2), the repeatability of measurement for this dimension is deemed compliant.

It is observed that the repeatability for dimension N°2 part 3 gives the maximum value of 10% of the total tolerance. As this is below the minimum standard established for "Major" feature category (Repeatability <20% from Table 2), the repeatability of measurement for this dimension is deemed compliant.

The measurement system and measurement program has acceptable levels of repeatability so the study records shall be maintained internally and archived properly.

If several operators or fixtures for positioning are introduced, a further gauge repeatability and reproducibility study will be required to establish if they affect the measurement system. This simple test has not tested the measurement system accuracy so further tests may be advised.

10.4 Case Study – Gauge R&R

An aero engine supplier is manufacturing machined structures. An inspection device is used to determine a critical feature on one of the parts. To evaluate the measurement system and determine if it is fit for its intended purpose a MSA is conducted.

The critical feature is an outer diameter with specification limits 838.60 - 838.80 mm (total tolerance = 0.2 mm). The inspection device is a dial gauge comparator together with a master gauge.

The manufacturing engineer started planning the MSA study by considering potential sources of variation in the measurement system and developing the test procedures, making sure the test followed the defined measuring procedure and evaluating if any environmental factors affect the measurement system.

The gauge resolution and the accuracy ratio are determined to meet the acceptance criteria. Since the measurement system includes several operators, the engineer decides to perform a Gauge R&R study. Gauge R&R can be generated using a number of statistical software packages but can also be calculated manually.

Ten parts were selected that represent the expected range of the process variation. Three (3) operators measured the ten parts, three (3) times per part, in a random order without seeing each other's readings.

Data:

	Operator A			Operator B			Operator C		
Part	**Trial 1**	**Trial 2**	**Trial 3**	**Trial 1**	**Trial 2**	**Trial 3**	**Trial 1**	**Trial 2**	**Trial 3**
1	838.79	838.77	838.80	838.78	838.77	838.79	838.78	838.80	838.79
2	838.69	838.68	838.70	838.69	838.70	838.72	838.72	838.69	838.73
3	838.72	838.69	838.71	838.70	838.71	838.73	838.72	838.74	838.71
4	838.75	838.74	838.73	838.73	838.75	838.73	838.76	838.76	838.72
5	838.73	838.72	838.70	838.71	838.73	838.72	838.73	838.73	838.75
6	838.77	838.79	838.79	838.77	838.79	838.77	838.78	838.79	838.78
7	838.67	838.68	838.69	838.70	838.69	838.66	838.68	838.67	838.70
8	838.60	838.61	838.62	838.61	838.64	838.60	838.62	838.60	838.62
9	838.63	838.65	838.66	838.66	838.63	838.65	838.66	838.65	838.64
10	838.78	838.78	838.77	838.77	838.78	838.75	838.77	838.75	838.76

The mean average diameter for each part/operator combination was then calculated together with the range (maximum - minimum).

Operator A		
Part	**Mean Average**	**Range**
1	838.787	0.030
2	838.690	0.020
3	838.707	0.030
4	838.740	0.020
5	838.717	0.030
6	838.783	0.020
7	838.680	0.020
8	838.610	0.020
9	838.647	0.030
10	838.777	0.010
Overall averages	838.7137	0.0230

Operator B		
Part	**Mean Average**	**Range**
1	838.780	0.020
2	838.703	0.030
3	838.713	0.030
4	838.737	0.020
5	838.720	0.020
6	838.777	0.020
7	838.683	0.040
8	838.617	0.040
9	838.647	0.030
10	838.767	0.030
Overall averages	838.7143	0.0280

Operator C		
Part	**Mean Average**	**Range**
1	838.790	0.020
2	838.713	0.040
3	838.723	0.030
4	838.747	0.040
5	838.737	0.020
6	838.783	0.010
7	838.683	0.030
8	838.613	0.020
9	838.650	0.020
10	838.760	0.020
Overall averages	838.7200	0.0250

Calculating all the variance components from the above data gave us the following results, using the simpler Xbar R method:

Gauge R&R for Diameter:

Source	*Variance Component*	*% Contribution (of Variance Component)*
Total gauge R&R	0.0002274	7.20
Repeatability	0.0002239	7.08
Reproducibility	0.0000035	0.11
Part-to-part	0.0029331	92.80
Total variation	0.0031605	100.00

Source	*Standard Deviation (SD)*	*Study Variation (6 * SD)*	*% Study Variation*	*% Tolerance (SV/Tolerance)*
Total gauge R&R	0.0150811	0.090487	26.83	45.24
Repeatability	0.0149636	0.089781	26.62	44.89
Reproducibility	0.0018792	0.011275	3.34	5.64
Part-to-part	0.0541579	0.324948	96.33	162.47
Total variation	0.0562185	0.337311	100.00	168.66

Number of distinct categories = 5

This showed that the total gauge R&R value when calculated as a percent of the tolerance was 45.24%, more than the permitted value of 10% maximum allowed for critical features, so the measurement system was rejected as being not suitable to determine the diameter of the structure.

Recalculating the results using the recommended ANOVA method gave the following results:

Two-way ANOVA Table without Interaction:

Source	***Degrees of Freedom***	***Sum of Squares***	***Mean Squares***	***F statistic***	***p-value***
Part	9	0.260893	0.0289881	164.562	0.000
Operator	2	0.000727	0.0003633	2.063	0.134
Repeatability	78	0.013740	0.0001762		
Total	89	0.275360			

Gauge R&R for Diameter:

Source	***Variance Component***	***% Contribution (of Variance Component)***
Total gauge R&R	0.0001824	5.39
Repeatability	0.0001762	5.21
Reproducibility	0.0000062	0.18
Operator	0.0000062	0.18
Part-to-part	0.0032013	94.61
Total variation	0.0033837	100.00

Source	***Standard Deviation (SD)***	***Study Variation (6 * SD)***	***% Study Variation***	***% Tolerance (SV/Tolerance)***
Total gauge R&R	0.0135053	0.081032	23.22	40.52
Repeatability	0.0132723	0.079634	22.82	39.82
Reproducibility	0.0024979	0.014987	4.29	7.49
Operator	0.0024979	0.014987	4.29	7.49
Part-to-part	0.0565803	0.339482	97.27	169.74
Total variation	0.0581698	0.349019	100.00	174.51

In this example the Gauge R&R % Tolerance figure was 40.52%, slightly lower than the Xbar R method but still greater than the permitted maximum of 10%.

NOTE: The calculations used to determine these results are available in the AIAG guide 'Measurement System Analysis' referenced in Section 2.

10.5 Case Study – Gauge R&R for Co-ordinate Measurement Machine

See also CMM program verification case study in 10.1.

Co-ordinate measuring systems can come in a number of configurations and should be considered as a system of measuring co-ordinate data from a datum point. This might include tactile systems such as coordinate measuring machine touch trigger probing, or structured light measurement vision systems that gather images of the component and then establish the coordinate system within software. In all cases it is important to understand the measurement process used and establish the tests required to ensure the system results are capable for the measurements being conducted. Typically coordinate measuring systems will use the following steps to establish a measurement:

1. System qualification – Establish a measurement reference system by qualification of styli tips, axes or vision areas.

2. Establish the reference or datum systems for the component

3. Measure the component features required from the datum points

4. Determine a measurement value through the calculation of features based on the coordinate data. This may be based on the fitting of mathematically perfect features to the measured data, or by comparison of the data against a known standard such as solid model.

5. Report the measured results.

   NOTE: More information on coordinate measurement system causes of variation are listed in 10.10 of this report.

While coordinate measuring systems are generally thought of as accurate and repeatable, the influence of the system, environment, calibration, program, drawing interpretation, system qualification fixturing, etc., can influence measurement capability. In all instances it is important to establish:

1. Measurement repeatability (the measurement system itself)

2. Measurement reproducibility (includes all the variables of measuring the part)

3. Measurement accuracy or bias (required to check the CMM program is correct)

Design of Experiments:

Design of Experiments (DoE) is a subject in its own right but the basic idea is simple and lies at the heart of the Gauge R&R method. The principle is that the measurement system will be affected by a number of process inputs to produce a single measured result as a system output. The purpose of DoE is to understand how the output varies with different combinations of the inputs.

A Gauge R&R study is an experiment designed to understand the variation in the output of the measurement system when the measurement system is subject to external sources of variation. The key to a good Gauge R&R study on a CMM is sound experimental design.

If a measurement system is used to repeat the same measurement several times whilst nothing is altered, any variation in the measured data is attributable to variation inherent in the measuring equipment itself. Repeating the measurement whilst keeping everything constant should, in theory, give an answer which is equal to the CMM accuracy and pure CMM repeatability. Any observed variation can only be due to the measurement equipment itself because all other possible influence factors have been held constant. Likewise, reproducibility tests should be designed to include variations that will be seen in the measurement process.

Sources of Variation:

Operator influence tends to dominate hand held gauging and so variation between operators is usually the only reproducibility factor tested for in a conventional gauge study. Conversely, a co-ordinate measuring system executing a part program will not necessarily be influenced by the person who presses the buttons, but will be influenced by other factors. However, as a co-ordinate measuring system operates at a much higher level of precision than hand held gauges, there will be many other sources of variation which will have an effect and will therefore have to be accounted for in the design of the Gauge R&R study. Possible sources of variation affecting coordinate measuring systems include:

- Variation in ambient temperature (measurement system growth)
- Variation in component temperature (part growth)
- Choice of probing stylus (long styli increase errors)
- Condition of the stylus (degree of wear, damage, or cleanliness)
- Location of the component in the machine measurement volume
- Variation in the location of the part in any fixturing used on the measurement table. Variation may induce the probe stylus to hit the part before the measurement sphere
- CMM to CMM variation if two or more machines are used in parallel on the same part
- Variation between part programs (scanning, touch trigger, number of points, position of points measured on the component, mathematics, component drawing interpretation, etc.)
- Variation in surface finish or part size may cause the measurement to be taken at a different point on the component
- Variation in fixture size
- Operator influence through variation in set up, probe qualification, datuming practice or assessment of probe qualification results
- Condition and cleanliness of machine, component, etc.

Sound CMM stewardship may standardize many of the above and minimize some of the sources of variation to rendered them insignificant. Nevertheless, it is likely that in most situations there will be some reproducibility factors that will influence the measurement.

The Design of the Study:

Gauge R&R studies should consider every combination of influencing factors that can be anticipated. The start point is to obtain a set of components that represent around 80% or more of the range of output from the manufacturing process. The second step is to decide what reproducibility factors need to be tested for.

Suppose the chosen factors are variation in ambient temperature and machine to machine variation. The ambient temperature variation might conveniently be characterized as ‘morning’ or ‘afternoon’ (i.e., two levels) or it might be an actual temperature reading taken in the measurement area. Care is needed here because the analysis is different depending on whether temperature is a ‘fixed factor’ (i.e., two levels) or a ‘random factor’ (actual values recorded). The machine to machine influence can be conveniently characterized as ‘machine A’ and ‘machine B’. The experimental design will resemble something like that depicted in the table below.

Component	Machine A		Machine B	
Part 1	M/c A, am, run 1	M/c A, pm, run 1	M/c B, am, run 1	M/c B, pm, run 1
Part 2	M/c A, am, run 1	M/c A, pm, run 1	M/c B, am, run 1	M/c B, pm, run 1
Part 3	M/c A, am, run 1	M/c A, pm, run 1	M/c B, am, run 1	M/c B, pm, run 1

Part 1	M/c A, am, run 2	M/c A, pm, run 2	M/c B, am, run 2	M/c B, pm, run 2
Part 2	M/c A, am, run 2	M/c A, pm, run 2	M/c B, am, run 2	M/c B, pm, run 2
Part 3	M/c A, am, run 2	M/c A, pm, run 2	M/c B, am, run 2	M/c B, pm, run 2

The above table is a full factorial experimental design in which the variables are:

1. Run to run variation on the same machine,

2. Variation between machines

3. Variation between the time of day – taken as a surrogate for ambient temperature.

This data set is best analyzed using the ANOVA method. The overall variation is the total Gauge R&R. The 'run to run' contribution to the variation is the repeatability; the sum of the other contributions is the reproducibility.

This is an example of how variation in process inputs can be analyzed. It should be noted that each CMM run will inspect multiple features, and therefore the repeatability, reproducibility and Gauge R&R values will be calculated for every feature.

Conclusions:

A successful Gauge R&R study on a co-ordinate measurement system requires some initial insight into the likely sources of variation. An experiment must then be designed to test for the chosen sources of variation. An understanding of the relative contributions of the sources of variation will provide a good starting point and can be made generic through control of the process. If a good level of control is implemented, many causes of variation will be limited and can be excluded from future studies. Where very tight component tolerances are being measured, more causes of variation will have an influence and should be included in the study.

10.6 Case Study – Attribute

SITUATION: An assembly shop had a high number of escapes due to lockwire issues. No matter how much training they conducted, escapes continued to be a problem. The engineer responsible for the product decided to run a Kappa evaluation to see if the operators installing the lockwire, and the inspectors inspecting the work were able to recognize the difference between acceptable and unacceptable. Twenty samples were collected, ten acceptable and ten unacceptable. The engineer suspected that the problem was not with the obvious conditions but with those on the edge of the acceptance criteria. So within each group they selected samples where one was an obvious example of that group while the remaining nine were just on the edge of the acceptance criteria. The engineer selected two operators who installed the lockwire and two inspectors who final inspected the parts as evaluators for the study.

In order to maintain an independent evaluation, each evaluator was only allowed to view one part at a time while the balance of the samples was out of view. Each evaluator performed their evaluation out of sight from the other evaluators and the engineer performing the evaluation randomized the order of sample presentation. After each evaluator had inspected every sample twice the engineer analyzed the results. First they analyzed the agreement within each evaluator to themselves and then they analyzed the agreement between evaluators using the first inspection from each evaluator. As expected the obvious samples were evaluated correctly every time. The differences were on the parts marginally acceptable or unacceptable.

Base line		Actual run for Evaluator A (randomized)				Evaluator A			Evaluator B			Evaluator C		
Sample	Truth	Evaluator	Sample	Insp. 1	Insp. 2	Sorted	Insp. 1	Insp. 2	Sorted	Insp. 1	Insp. 2	Sorted	Insp 1	Insp. 2
1	Good	A	20	Bad	Bad	1	Good	Good	1	Good	Good	1	Good	Good
2	Good	A	3	Good	Good	2	Good	Good	2	Bad	Good	2	Good	Bad
3	Good	A	1	Good	Good	3	Good	Good	3	Good	Good	3	Good	Good
4	Good	A	6	Good	Good	4	Good	Bad	4	Good	Bad	4	Good	Good
5	Good	A	18	Bad	Bad	5	Bad	Bad	5	Good	Good	5	Bad	Bad
6	Good	A	9	Good	Good	6	Good	Good	6	Bad	Good	6	Good	Good
7	Good	A	12	Bad	Good	7	Good	Good	7	Good	Good	7	Good	Good
8	Good	A	14	Bad	Bad	8	Bad	Good	8	Good	Bad	8	Good	Bad
9	Good	A	15	Bad	Bad	9	Good	Good	9	Good	Good	9	Good	Good
10	Good	A	4	Good	Bad	10	Good	Bad	10	Good	Good	10	Bad	Good
11	Bad	A	11	Bad	Bad	11	Bad	Bad	11	Bad	Bad	11	Bad	Bad
12	Bad	A	7	Good	Good	12	Bad	Good	12	Good	Bad	12	Bad	Bad
13	Bad	A	17	Good	Bad	13	Good	Good	13	Bad	Bad	13	Good	Good
14	Bad	A	13	Good	Good	14	Bad	Bad	14	Bad	Good	14	Bad	Bad
15	Bad	A	10	Good	Bad	15	Bad	Bad	15	Good	Bad	15	Bad	Bad
16	Bad	A	16	Bad	Bad	16	Bad	Bad	16	Bad	Bad	16	Bad	Bad
17	Bad	A	8	Bad	Good	17	Good	Bad	17	Bad	Good	17	Good	Good
18	Bad	A	5	Bad	Bad	18	Bad	Bad	18	Bad	Bad	18	Bad	Bad
19	Bad	A	19	Bad	Good	19	Bad	Good	19	Good	Bad	19	Bad	Bad
20	Bad	A	2	Good	Good	20	Bad	Bad	20	Bad	Bad	20	Bad	Bad

Evaluator A					Kappa Score	
		Insp. 1			P (o)	0.700
	20	Good	Bad		P (c)	0.500
Insp. 2	Good	7	3	0.500	**K =**	**0.400**
	Bad	3	7	0.500		
		0.500	0.500			

Evaluator B					Kappa Score	
		Insp. 1			P (o)	0.550
	20	Good	Bad		P (c)	0.500
Insp. 2	Good	6	4	0.500	**K =**	**0.100**
	Bad	5	5	0.500		
		0.550	0.450			

Evaluator C					Kappa Score	
		Insp. 1			P (o)	0.850
	20	Good	Bad		P (c)	0.500
Insp. 2	Good	8	1	0.450	**K =**	**0.700**
	Bad	2	9	0.550		
		0.500	0.500			

Evaluator A Insp. 1 to B Insp. 1			Evaluator A Insp. 1 to C Insp. 1			Evaluator B Insp. 1 to C Insp. 1			C Insp. 1 vs Truth		
Sorted	Insp. 1	Insp. 2	Sorted	Insp. 1	Insp. 2	Sorted	Insp. 1	Insp. 2	Sorted	Insp. 1	Truth
1	Good	Good	1	Good	Good	1	Good	Good	1	Good	Good
2	Good	Bad	2	Good	Good	2	Bad	Good	2	Good	Good
3	Good	Good	3	Good	Good	3	Good	Good	3	Good	Good
4	Good	Good	4	Good	Good	4	Good	Good	4	Good	Good
5	Bad	Good	5	Bad	Bad	5	Good	Bad	5	Bad	Good
6	Good	Bad	6	Good	Good	6	Bad	Good	6	Good	Good
7	Good	Good	7	Good	Good	7	Good	Good	7	Good	Good
8	Bad	Good	8	Bad	Good	8	Good	Good	8	Good	Good
9	Good	Good	9	Good	Good	9	Good	Good	9	Good	Good
10	Good	Good	10	Good	Bad	10	Good	Bad	10	Bad	Good
11	Bad	Bad	11	Bad	Bad	11	Bad	Bad	11	Bad	Bad
12	Bad	Good	12	Bad	Bad	12	Good	Bad	12	Bad	Bad
13	Good	Bad	13	Good	Good	13	Bad	Good	13	Good	Bad
14	Bad	Bad	14	Bad	Bad	14	Bad	Bad	14	Bad	Bad
15	Bad	Good	15	Bad	Bad	15	Good	Bad	15	Bad	Bad
16	Bad	Bad	16	Bad	Bad	16	Bad	Bad	16	Bad	Bad
17	Good	Bad	17	Good	Good	17	Bad	Good	17	Good	Bad
18	Bad	Bad	18	Bad	Bad	18	Bad	Bad	18	Bad	Bad
19	Bad	Good	19	Bad	Bad	19	Good	Bad	19	Bad	Bad
20	Bad	Bad	20	Bad	Bad	20	Bad	Bad	20	Bad	Bad

Evaluator A Insp. 1 to B Insp. 1					Kappa Score	
		Insp. 1			P (o)	0.550
	20	Good	Bad		P (c)	0.500
Insp. 2	Good	6	5	0.550	**K =**	**0.100**
	Bad	4	5	0.450		
		0.500	0.500			

Evaluator A Insp. 1 to C Insp. 1					Kappa Score	
		Insp. 1			P (o)	0.900
	20	Good	Bad		P (c)	0.500
Insp. 2	Good	9	1	0.500	**K =**	**0.800**
	Bad	1	9	0.500		
		0.500	0.500			

Evaluator B Insp. 1 to C Insp. 1					Kappa Score	
		Insp. 1			P (o)	0.550
	20	Good	Bad		P (c)	0.500
Insp. 2	Good	6	4	0.500	**K =**	**0.100**
	Bad	5	5	0.500		
		0.550	0.450			

C first measure vs Truth					Kappa Score	
		Insp. 1			P (o)	0.800
	20	Good	Bad		P (c)	0.500
Insp. 2	Good	8	2	0.500	**K =**	**0.600**
	Bad	2	8	0.500		
		0.500	0.500			

Kappa Formula

$$K = \frac{\text{P observed - P chance}}{\text{1 - P chance}}$$

How to fill out the contengency table

Both agree Good	First says Bad second says Good
First says Good second says Bad	Both agree Bad

Math Explained 20 pc. Sample

Convert observed values into decimal point percent*

9	2
1	8

Becomes

0.450	0.100
0.050	0.400

Then total columns and rows

0.450	0.100	0.550
0.050	0.400	0.450
0.500	0.500	

P observed is the sum of agreement

0.450	0.100
0.050	0.400

P observed = .450 + .400 = .850

P chance is the probability of each classification mulitplied and then summed

0.450	0.100	0.550
0.050	0.400	0.450
0.500	0.500	

P chance = (.500 x .550) + (.500 x .450) = .500

Kappa for this example is

$$K = .700 \quad \frac{\text{P (o) - P (c) \{.850 - .500\}}}{\text{1 - P (c) \{1- .500\}}}$$

* You can sum the columns and rows first then convert to decimal and get the same results

The comparison between Evaluator A and C appear to show a capable system however, based on Evaluator A's results within themselves it is easy to discount this level of agreement was by chance not actual agreement. Evaluator C was the closest to an acceptable Kappa score when measured within them. However, the engineer noticed that two of the responses indicated the acceptance of bad product as good for both evaluations.

The engineer decided to compare evaluator C against the truth. The results showed a poor Kappa score of K= 6.

Results:

From the results of the Kappa analysis the engineer recognized that the training provided was not enough to ensure conforming product was built and shipped. They developed a reference board that demonstrated the subtle difference between acceptable and unacceptable of various possible conditions. All of the operators and inspectors were provided this reference board and re-trained using the reference board in the training. The engineer re-ran the Kappa analysis allowing the evaluators to use the reference boards. With the retraining and use of the reference boards all of the Kappa scores were acceptable with a few showing perfect agreement (score 1.0).

Even after the engineer had implemented the improved process they continued to have escapes until they provided their purchaser a reference board and training. It turned out that the purchaser's inspectors also required calibration.

Key Learning Point:

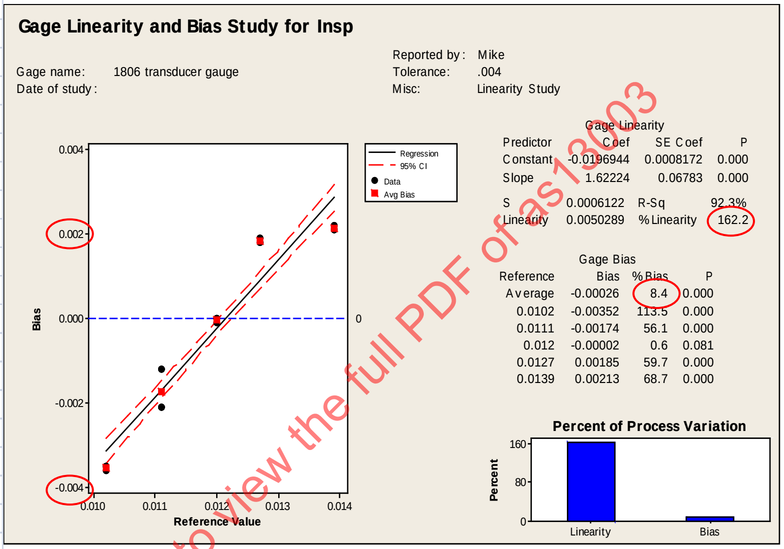
Training is not always enough to drive consistency in subjective measurement systems. Just because an evaluator is consistent does not mean their analysis is correct. The evaluator must utilize all the information collected to make the right decisions on how to move forward.

10.7 Case Study – Linearity

A casting supplier is asked to perform a MSA and Linearity/Bias study for all inspection equipment that will be used on a new turbine blade they are developing. One of the pieces of equipment is a transducer gauge which inspects 80% of the required features. The MSA study shows the gauge results well within the acceptance criteria for all features inspected, however, the Linearity/Bias study shows a Linearity problem on the feature where the gauge touches the air foil.

A review of the gauge showed the angle at which the transducer touches the part is based upon a nominal blade. A review of the blades used showed varying material condition and geometry on the datum locators and the inspection point. A review of the process capability over 3 production lots showed the feature nominally shifted. The gauge R&R results indicated the process variation was 0.0031.

Blade	Actual value	Insp
1	0.0102	0.0067
1	0.0102	0.0067
1	0.0102	0.0066
1	0.0102	0.0067
1	0.0102	0.0067
1	0.0102	0.0067
1	0.0102	0.0066
1	0.0102	0.0067
1	0.0102	0.0067
1	0.0102	0.0067
2	0.0111	0.0090
2	0.0111	0.0090
2	0.0111	0.0099
2	0.0111	0.0099
2	0.0111	0.0090
2	0.0111	0.0099
2	0.0111	0.0090
2	0.0111	0.0090
2	0.0111	0.0090
2	0.0111	0.0099



Initial Data:

As noted, while the results are repeatable, there is significant accuracy error as the part deviates from nominal. This error is impacting the observed process variation.

Results:

The dimensional and geometric variation was causing the transducer to touch the airfoil in different locations for each blade. The supplier experimented with the transducer angle until they found the angle least susceptible to material and geometry influences. Once the adjustments to the gauge were complete, the supplier re-ran the gauge R&R and the Linearity/Bias studies. Results observed in the gauge R&R where the process variation went down to 0.0015. In the Linearity/Bias study, while it did not totally eliminate the linearity error it reduced it such that it had no significant impact on the results.