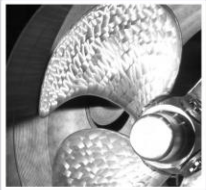
Measurement System Analysis

How-to Guide

Version 6.1
August 2013



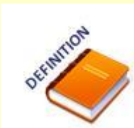
Contents

Rolls-Royce

How to Use This Guide

- This ‘How to Guide’ is designed as a complete training package that you can work through individually at your own pace (or in small teams as part of a facilitated training exercise).
- By carefully reading the text, and practising the tools in the associated ‘Workbook’, you will become competent and confident in using these process control tools in your work area.
- The How to Guide is designed to be applicable for use primarily by Manufacturing Engineers and Lean Sigma Practitioners – from any area of the business. For this reason, the technical explanations are based on general business application examples – to ensure everyone can relate to them. Throughout the guide, there are case study examples which show how the theory is applied at the different stages of the process control sequence.
- Before you start, make sure you also have the Workbook available. It is essential that you work through this in parallel with the How to Guide, and that you complete the practise questions, plus case study exercise before you start to use MSA techniques in the business.

Icons are used throughout to highlight key elements, and to signpost supplementary information where appropriate.

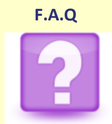


The technical explanation of the core terminology

Indicating where you must seek help from practitioner experts such as Black or Green Belts

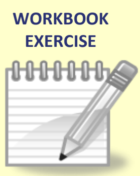
Where you can find additional information, and the next phase of the improvement journey

Indicates a key learning point



The approved Rolls-Royce answers to main queries asked by users

Tips on the commonly observed pitfalls - & how to avoid them

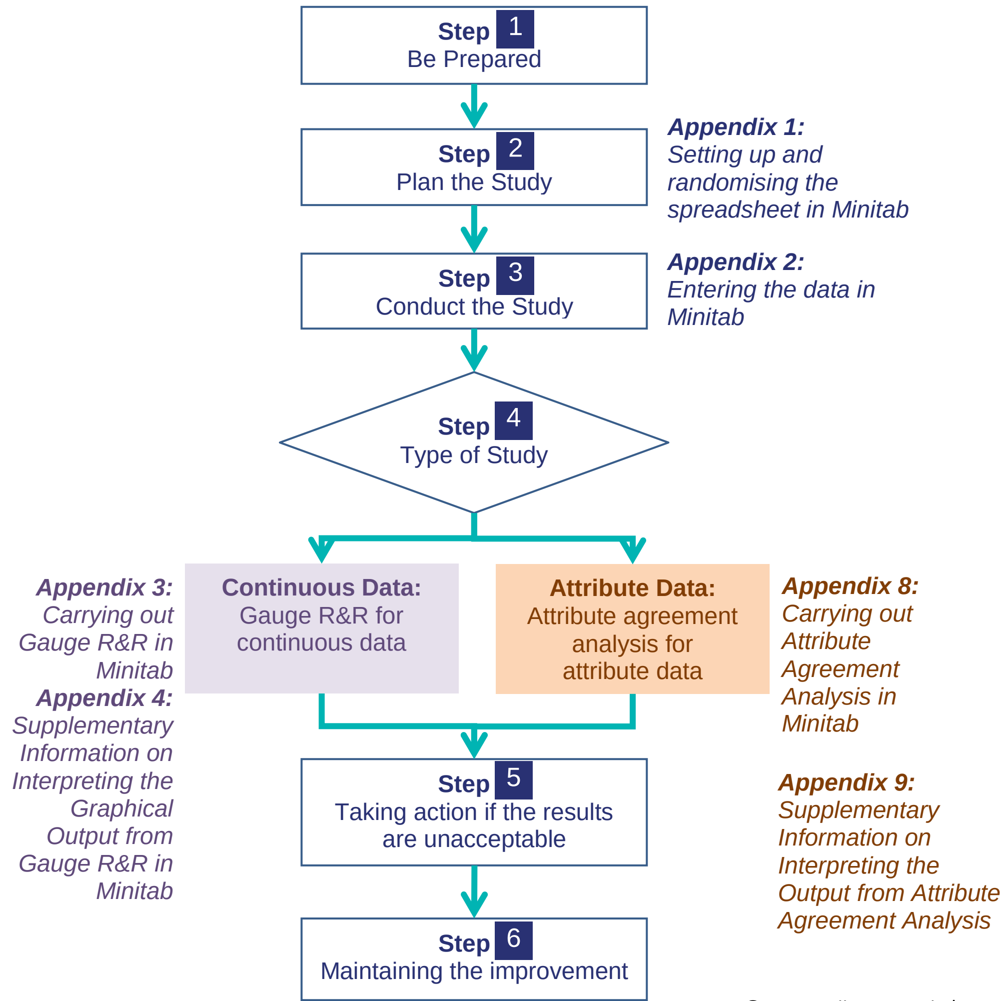


The separate Workbook which you must use in parallel with your learning

Guide Structure

The flowchart illustrates the structure of this Guide which is designed to provide step by step guidance for conducting and interpreting the Measurement System Analysis. Steps 1, 2, 3, 5 & 6 below are common for all types of MSA study. Step 4 differs depending on whether the study is for the collection of continuous or attribute measurements (see page 14 for full guidelines on how to select the appropriate type of study).

Supplementary information on how to carry out the analysis using Minitab statistical software – together with some of the more detailed analysis of the statistical output – is provided in the appendices as outlined below.



Section 1

Introduction

In this section:

1.1: The Need for Measurement System Analysis

In our day to day work we often collect and use data to make important decisions about our processes. For example we measure the critical dimensions of the parts we produce to check that they are correctly manufactured; we inspect documentation and drawings to check that they have been correctly completed; we test our engines to confirm that they are functioning to specification.

Think for a moment about the data that is collected in your own work area. Regardless of whether you work in a manufacturing, design or transactional function you will undoubtedly be able to think of examples of data which are regularly collected to confirm the quality of work, to monitor performance against targets or to allow in-process decisions to be made.

Now ask yourself can you trust that data? Are you sure it is reliable? Are you sure that the data is measured consistently? If more than one person or piece of measuring or test equipment is involved then are you sure that if each were to measure/inspect the same item that they would all reach the same conclusion?

If your answer to any of the above questions is ‘no’ then there is a real possibility that your measurement system could be producing unreliable data. This could lead your team to draw the wrong conclusions about whether your processes are in control and capable. Unreliable data can lead us to believe there is a problem with the process when actually everything is OK or it may prevent us from spotting a problem. This is likely to cost the business money, through unnecessary scrap or rework, or unnecessary improvement projects, or through being unaware that there is a problem with a process or product, and risking customer complaints or more serious problems such as safety incidents.

Therefore, before collecting and using data to make decisions about any process or product, it is important to check that the measurement system is good enough by doing a Measurement System Analysis.

1.2: What is Measurement System Analysis?

Measurement System Analysis (usually referred to as **MSA**) is a structured procedure which we use to assess the ability of a measurement system to provide good quality data.

All data is collected using some form of **Measurement System**. A measurement system is any system used to collect data. For example it can be a piece of equipment (such as a clock), some software, or it could be a person (such as an inspector recording the count of defects in a batch of materials) or a combination of these.

A **Measurement System** is the combination of people, equipment, materials, methods and environment involved in obtaining measurements

Examples:
Data: time taken from receipt of an invoice until payment
Measurement System: a piece of computer software which measures the time from start to finish, or possibly someone measuring the time using a stopwatch or a clock

Data: Dimensions of a machined part
Measurement System: a vernier calliper, a coordinate measurement machine (CMM) or a Go/No Go gauge in combination with the person operating the piece of equipment, the procedures they are following and environmental conditions such as the temperature or vibration level in the working environment

Data: Number of errors on an engineering drawing
Measurement System: a person checking each drawing and counting the number of errors in combination with the assessment standards, procedures being used and environment conditions such as changes in the natural light conditions.

Potential Pitfall: Where the measurement system is a device such as a vernier calliper or CMM and you have multiple devices it is important to recognise that each device is a separate system and each must separately be analysed. For example just because one CMM machine is a good or bad system does not mean that other 'identical' machines will have the same characteristics.

WORKBOOK EXERCISE

Turn to **Exercise 1** in the Workbook to think about data and the associated measurement systems within your own working area.

1.3: Different Types of Measurement System Analysis

Although the principles are similar, there are 2 main types of Measurement System Analysis (MSA). The appropriate one to use depends on the type of data being collected using the measurement system.

Where the measurement system is collecting data which is measured on a continuous scale such as time, weight, dimensions or pressure then we will be collecting and comparing numerical results (such as the weights of parcels in grams). This type of data is called **continuous data.** In this case we use an MSA method called **Gauge R&R**

Continuous data comes from measurements on a continuous scale such as: temperature, time, distance, weight, dimensions.

Gauge R&R is short for Gauge Repeatability and Reproducibility. It is the Measurement System Analysis method used to analyse measurement systems which collect continuous data

Where the measurement system collects data which categorises each item (such as pass/fail, grading items into different categories or counting defects such as scratches in paintwork or errors on a drawing) then we will be analysing agreement in the categories chosen. This type of data is Called **Attribute data.** In this case we use an MSA method called **Attribute Agreement Analysis**

Attribute data is based on upon counting how many units fall into discrete distinctions such as: pass/fail or percentage defective

Attribute Agreement Analysis is the Measurement System Analysis method used to analyse measurement systems for attribute data

WORKBOOK EXERCISE

Turn to **Exercise 2** in the Workbook to check your understanding of the difference between continuous and attribute data

1.4: How does Measurement System Analysis work?

Regardless of the type of MSA being used, there are two key things we are usually interested in when we analyse a measurement system

1) If the ***same*** person or piece of equipment measures the ***same*** item over and over again do we consistently get the same data? For example if the same person measures the weight of a parcel 3 times would they record the same weight each time? If a team member assesses whether an expenses claim should be paid or rejected and looks at the same claim 3 times would they be consistent in their decision?

This is important to know. If a person or a piece of equipment is not able to consistently measure the same item then clearly this would result in unreliable data.

This check for consistency **within** an inspector is referred to using slightly different terminology depending on the type of measurement being made.

For measurement of continuous data such as dimensions, weight or time then this check for consistency is referred to as **Repeatability**

Repeatability assesses whether each person can measure the same item multiple times with the same measurement device and get the same value when measuring **continuous data**.

For measurement of attribute data such as the invoice checking example above then this check for consistency is referred to as **Within Appraiser Agreement**

Within Appraiser Agreement assesses whether each person can assess the same item multiple times using the same measurement method and get the same result when measuring **attribute data**

1.4: How does Measurement System Analysis work? (cont.)

2) If ***different*** people or pieces of equipment measure the same item would they each get the same result? For example if three different people weighed the same parcel multiple times would they each record the same weight? If every member of the accounts team were asked to assess whether they would pay or reject an expenses claim would they all draw the same conclusion?

Again this is important to understand. Problems such as 'no fault found' part returns are often a result of a lack of agreement between two different measurement systems. Obviously any disagreement indicates a problem with the reliability of data collected from such a system

This check for consistency **between** inspectors is also referred to using slightly different terminology depending on the type of measurement being made.

For measurement of continuous data such as dimensions, weight or time then this check for consistency between inspectors is referred to as **Reproducibility**

Reproducibility assesses whether different people can measure the same item multiple times with the same measurement device and get the same average value

For measurement of attribute data such as the invoice checking example above then this check for consistency is referred to as **Between Appraiser Agreement**

Between Appraiser Agreement assesses whether different people can assess the same item using the same measurement method and get the same result when measuring **attribute data**

WORKBOOK EXERCISE

Turn to **Exercise 3** in the Workbook to check your understanding of these definitions.

1.4: How does Measurement System Analysis work? (cont.)

Measurement System Analysis works by setting up and running a controlled experiment to check the **repeatability** and **reproducibility** or the within appraiser and between appraiser agreement of the system

Essentially a number of items (such as machined parts or drawings for inspection) are selected and deliberately measured multiple times by each person or piece of equipment. We then analyse the results using Minitab software (full details on how to do this are provided in the appendices), to establish how consistent the measurement system is, and to decide whether improvements need to be made before using it to collect data.

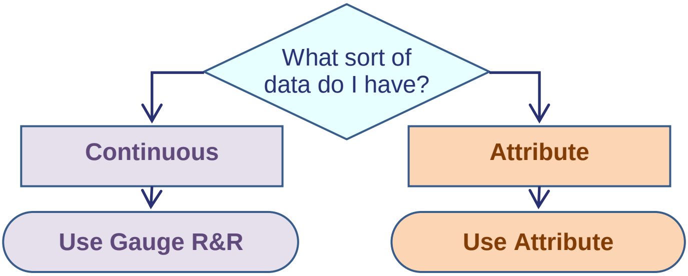
For example, to run such an experiment in the parcel weighing example you would select a range of parcels (around 10 would be sufficient in this case) and get a number of different people who work in the area to each weigh the parcels a number of times. As we will explain in section 2 usually three different people are selected to make the measurements and each item is measured by each person three times. This provides sufficient data to be able to check for consistency both within and between those people making the measurements.

If a problem is found with the measurement system then the next step is to find the root cause of the problem and try to eliminate as much variation as possible within the measurement system. In the parcel weighing example differences between operators may exist due to them having different methods for using the scales or due to them interpreting or recording the weight display in different ways.

Teams are often surprised to discover that measurement systems they have used for years are actually inconsistent. Carrying out formal Measurement System Analysis is a good structured way for teams to assess and improve their measurement system.

You may well already make informal checks of this kind. However using the formal structured approach provided by Measurement System Analysis allows us to be consistent and thorough in the way that we check the reliability of our data.

1.5: Choosing the Appropriate Type of Study



The flow chart above can be used to select the types of measurement study to conduct

SEEK GUIDANCE

Help

If you are not sure what type of data your variable is then ask a local Black Belt to help you choose the most appropriate type of MSA.

Summary

- A **Measurement System** is the combination of people, equipment, materials, methods and environment involved in obtaining measurements.
- If left unchecked there is a real risk that the measurement systems we use could be delivering unreliable data without us being aware of the existence of a problem. Therefore it is important to analyse how effective our critical measurement systems are.
- **Measurement System Analysis** is a structured procedure which we use to assess the ability of a measurement system to provide good quality data, before using it to collect data on which decisions are based.
- **Measurement System Analysis** provides a structured approach for teams to assess whether the measurement systems they use are fit for use.
- There are 2 main types of Measurement System Analysis:
  - **Gauge R&R**, which is the method used to evaluate measure systems for continuous data
  - **Attribute Agreement Analysis**, which is the method used to evaluate measurement systems for attribute data

WORKBOOK EXERCISE

Turn to **Exercise 4** in the Workbook to check your understanding of the key points so far.

Section 2

Conducting an MSA

In this section:

- A step-by-step guide for conducting an MSA

The key steps we will cover are as follows:

Step 1	**Be Prepared**
Step 2	**Planning the Study**
Step 3	**Conducting the Study**
Step 4	**Interpreting the Results for Continuous / Attribute Data**
Step 5	**Taking Action if the Results are Unacceptable**
Step 6	**Maintaining the Improvement**

Case Study Examples

We will introduce you to the application of these steps using two example scenarios which you can work through – one for a measurement system which collects continuous data and one for a system which collects attribute data. You will then have the opportunity to turn to the workbook and practise applying the tools yourself using further case study examples

Case Study - Scenario 1

Julie is the production manager of a factory where mobile phones are manufactured and assembled. There has recently been a noticeable increase in mobile phones being returned by customers because the touch screens aren't functioning correctly. She speaks to the failure analysis team, who examine every phone which is returned, and they have done some analysis which indicates that the screens are being manufactured with an incorrect thickness.

Julie sets up a team to investigate this – this is a major issue for the company, as every phone which is returned needs to be replaced at significant cost. She is also concerned that the reputation of the company's phones will be damaged, which may quickly result in loss of business to their competitors.

On initial investigation, the team can't find a reason for the screen thickness being incorrect. The thickness (in millimetres) of every 10th screen made is measured, using a vernier calliper, by one of the experienced measurement technicians. The data collected over the last few months indicates that there has been no issue with screen thickness.

Julie and her team can't understand why there seems to be a problem with screen thickness, yet all of their measurements show that the thickness is fine. One of the engineers suggests that perhaps there is an issue with the measurement of screen thickness – perhaps the data being collected is incorrect or unreliable in some way.

The measurement technicians don't believe there can be an issue with the measurements, as the measurement technique is so straightforward and they all have several years' experience in making the measurements. However, the team decides to carry out an MSA to eliminate this as a possible root cause.

Since the data being collected is continuous data (screen thickness measured in millimetres), the team recognise that they need to conduct a Gauge R&R study to determine repeatability and reproducibility.

Case Study Scenario 2

Tom works in the Quality department of the same major mobile phone manufacturer as Julie. There has been an increase in the number of mobile phones being rejected at the final test stage due to scratched screens. Tom cannot understand why this is happening, as each screen is inspected for scratches before assembly, and immediately has a protective film placed over it. This inspection is referred to as the in-line inspection, and there has been no increase in the number of scratches detected at the in-line inspection.

This issue results in around 5% of manufactured screens being scrapped. The screens are unable to be reworked adding a significant cost to the company. There is also an additional cost of fitting a replacement screen to each of the affected phones.

Tom feels that there might be an issue with the in-line inspection. The inspection is done by a team of quality technicians. They do a visual check of every screen manufactured, by placing it under an inspection lamp and looking for scratches. Each screen is then categorised as a Pass or a Fail and the results are recorded by the technicians in a database. Tom discusses the issue with the technicians, however they are adamant that there isn't a problem with the inspection as it has always worked well in the past. Nevertheless, Tom feels it would be worthwhile to conduct an MSA, so that inspection can be eliminated as a potential root cause.

The measurement system in this case consists of the person doing the inspection, plus the inspection lamp, methods, acceptance criteria being used and any environmental factors.

As the data being collected is either a Pass or a Fail, this is attribute data and therefore Attribute Agreement Analysis is the appropriate technique to determine the within and between appraiser agreement.

In both these measurement systems the parts are measured or inspected by people. Throughout this guide we will call these people the **appraisers.**

Appraisers: these are the people who measure the parts

Remember **repeatability** and **within appraiser agreement** are asking the question 'do the appraisers agree with themselves when asked to do the same assessment a number of times?' **Reproducibility** and **between appraiser agreement** are asking the question 'do the appraisers agree with each other when asked to do the same assessments?'

Step 1

Be Prepared

Ensure the pre-requisites are in place

We must remember that this activity builds on the process basics and it is important to ensure that these foundations are in place.

Prerequisites	Check
Equipment used is calibrated	
Equipment used has sufficient measurement resolution	
That the measurement system is suitable for the full range of measurements	
Equipment used is stable	
That there are clear operational definitions in place	
Observe a few measurements being made	

1a) Check equipment calibrated

There are then a few more prerequisites that we should check:
If the measurement system involves a piece of equipment, we must ensure that, where relevant, the calibration is up to date.

If the calibration is not up to date, there is a risk that there will be some **Bias** in the measurements, which means that the results will be different than their reference value. Bias is often also known as accuracy. An example of bias is if your bathroom scales are incorrectly set up (incorrectly calibrated) so that they consistently over-estimate your weight by a set amount such as 1kg.

For visual inspections it is also important to check that the appraisers have current eye examinations that meet the required standards.

The equipment having a current calibration sticker does not guarantee that it is measuring correctly. Bear in mind that the equipment may have been dropped, otherwise damaged or tampered with since it was last calibrated. Inspect the gauges carefully for any signs of damage and if in doubt contact your local Metrology department for advice

1b) Check the equipment resolution

Resolution is used to assess the increments of measure that a particular gauge can work to when measuring parts.

Where we are measuring **continuous data** it is important to check that measurement equipment has adequate resolution for the characteristic being measured.

For example, a set of bathroom scales are often marked in 1kg increments so would not have sufficient resolution for weighing baking ingredients if measurement was required to the nearest 10g.

As a guide, a suitable resolution for the measurement system is usually 1/10 of the tolerance or better. So for example if you needed to be able to measure baking ingredients out to be 100g +/- 5g then the tolerance would be 10g meaning that the required resolution of the measurement system should be 1g. In this case we would need to have scales which could measure to increments of 1g.

For attribute agreement ensure that **reference standards** are available and are able to distinguish between pass and fail conditions.

Reference standards provide clear guidance for appraisers to allow them to distinguish between pass and fail conditions. These could be written procedures, pictures or physical samples of parts which display both the pass and fail conditions.

1c) Check measurement system is suitable

For **continuous data** it is also important to ensure that the method being used is suitable for the full range of measurements you will be making – for example that it is equally suitable for measuring the smallest and largest parts. For example, when weighing airline baggage, bathroom scales would probably be suitable for weighing small holdalls, but would they be suitable (and give reliable results) for extra large suitcases?

1d) Check equipment is stable

We must also be confident that the measurement system is stable over time – that is, there is no reason to expect that the performance of the measurement system will vary over time. One way to check this is to measure the same part at regular intervals, and plot on an SPC chart.

For more information on SPC Charts see the 'How To Guide' for SPC Charts.

Continuous Data

Checking the equipment is calibrated, has suitable resolution and range and is stable over time

Julie and her team worked with their local Black Belt (Alison) to ensure that the pre-requisites are in place:

- The vernier calliper used to measure the screen thickness has been calibrated within the last year, which is fine.
- The specification limits for screen thickness are 1.5 – 2mm, so the tolerance is 0.5 mm. For the vernier calliper to have sufficient resolution, it must be able to measure in 0.05mm increments - it can actually measure down to 0.01mm, so meets this requirement easily.
- The team are satisfied that the measurement system is suitable for the full range of measurements.
- The team are also satisfied that the vernier calliper is stable – they have what they call a 'reference standard screen', which is stored in a cupboard and measured once per week (it is always measured by the same technician). The results from this standard screen have been very consistent (stable) over the past year (the data is plotted on an SPC chart so they can see this at a glance)

Attribute Data

Checking the equipment is calibrated, has suitable resolution and range and is stable over time

Tom carries out the following checks to ensure the pre-requisites are in place:

- Each appraiser has current eye examinations to the company standard.
- Reference standard screens are available, represent the full range of expected defects and have current calibration stickers.
- The reference standards are in good condition
- The inspection lamps have all got stickers confirming that their lux levels have been checked in the last 12 months

1e) Ensure clear operational definitions

Without an operational definition, there will almost certainly be misunderstandings and inconsistencies in the way that data is collected using the measurement system.

An **Operational Definition** is a clear description of:

- What to measure; ensuring that all aspects of the measure are carefully defined.
- How to measure it; including what equipment to use, how to use the equipment or standards and how to record the data.

Don't assume that it's obvious to everyone how to make the measurement!

Even measures that seem obvious are open to interpretation:

- For example, when making measurements using a ruler, does everyone start measuring at the point where zero is marked, or might some people start from the end of the ruler?
- If asking people to decide whether something is defective or not, does everyone have the same understanding of what 'defective' actually means? For example, if you were assessing the quality of the paintwork on a car door, some people might feel that a few tiny scratches is acceptable, whereas others may class the door as defective if it has any scratches at all.
- Ensure everyone has been trained in how to follow the operational definition

Turn to **Exercise 5** in the Workbook to practise writing operational definitions.

Case Study Continuous Scenario - Operational Definitions

Julie's team realise there is no operational definition for the measurement of the screen thickness, so they produce one:

- The screen should be held between the thumb and forefinger of one hand.
- Set the vernier calliper to zero
- The vernier calliper should be used to measure the thickness at a point in the middle of the screen.
- The thickness should be recorded, in millimetres, to two decimal places.

Continuous Data

Case Study Attribute Scenario - Operational Definitions

Tom sets up a team, including Alan, the Quality Technician team leader, to discuss running the Attribute Agreement Analysis. The team check that there is an operational definition in place for the in-line inspection, and in fact there is as follows:

- Review the reference standard screens
- Wearing gloves hold the bottom left hand corner of the screen using the silicon coated tweezers supplied.
- Do not touch any other part of the screen
- Hold the screen 5 centimetres below the inspection lamp.
- Tilt the screen in all directions – if any scratches are seen, contact the production engineer immediately.

The team feel that this operational definition is clear and they feel that they can proceed with the study.

Attribute Data

1f) Observe trial measurements

Finally, observe a few people making measurements, and note down any observations. Does it look like the operational definition is being followed? Is it confusing? Are there any misunderstandings or discrepancies?

If any issues are found during this observation, it is worth addressing these before moving on to conduct the MSA.

KEY POINT

When doing this observation, and also when conducting the study, it is important to clearly communicate to those involved that it is the Measurement System as a whole which is being assessed, not the individual people involved. To ensure that this is clear, people's names should not be used within the study, rather they should be referred to as 'Person A', 'Person B' etc, or something similar (ensuring that it is still possible to trace back to the real names where required). It is strongly advisable to think carefully about how to communicate the purpose of the MSA to the team. Where possible always try to communicate with the team face to face in a team meeting

Continuous Data

Initial Observations

There are 3 measurement technicians in Julie's team – Kevin, Mary and Karen – they will all participate in the study.

Julie observes each of them doing a few measurements, and although it appears that the 3 technicians are all following the same procedure, she feels that it is quite a difficult measurement to make. However, the technicians all say that they find it quite easy, so she decides to proceed with the Gauge R&R.

Attribute Data

Initial Observations

Over the next few days, Tom and Alan observe the technicians inspecting screens on a few occasions – as far as they can see they are following the operational definition.

Step 2

Planning the Study

2a) Selecting appropriate parts for the study
2b) Identifying appropriate people to involve in the study
2c) Deciding when to run the study
2d) Planning the data collection

2a) Selecting appropriate parts for the study

Once we are confident that the prerequisites are in place, we can begin to plan the MSA.

As we discussed in section 1, the principal of MSA is that we set up an experiment to measure several parts several times, and use statistical analysis (with the help of the Minitab software package) to look at how much variation there is in the results. We can then decide whether this amount of variation is acceptable, and if it is not, we can do further analysis to understand where the variation is coming from, and help us find the root cause(s).

The first step in planning an MSA is to consider which parts (or items) we are going to measure during the study. Remember that MSA can be applied to anything from manufactured parts to engineering drawings to invoices. For simplicity in this guide in all cases we will call the thing that is being measured the **part.**

Part: this refers to whatever is being measured or assessed. For example this could be a physical component which is being measured, a telephone call which is being timed or a document which is being checked for errors.

A common standard is to use a minimum of 10 parts for a Gauge R&R Study and minimum of 20 parts or samples for an Attribute Agreement Analysis. In both cases a standard approach is for each part to be measured by 3 different people, 3 times each – a total of 90 measurements.

Ideally, we want to select a range of parts – some from the lower end of the process range, some from the upper, and some in-between. Also include some borderline parts if possible (ones which are marginally inside or outside the required tolerance limits or standard). The reason for this is that we want to ensure that the amount of variation in the measurement system is the same across the whole range of parts, and if it isn't, we want to be able investigate why.

Potential Pitfall: Be careful not to confuse the process range with the customer specification. The process range is the full range of parts produced by the process. This may be wider or narrower than the tolerance specified by the customer. A failure to use the full process range will result in the measurement system analysis not being fully representative of the actual measurement system. If you have concerns about selecting a fully representative sample then contact your local Black Belt for advice.

SEEK GUIDANCE

If you are unsure as to which parts to choose for the study, or if it's not possible to obtain 10 parts, contact a local Black Belt for guidance.

Also seek guidance if it isn't possible to measure the same part more than once.

2b) Identify people to involve

The next step is to identify the appropriate people to involve in the study.

- First you need to choose the appraisers – the people who will be doing the repeat measurements of the parts. These should be people who normally make the measurements. The standard approach for MSA is to use 3 appraisers
- The 3 people should represent everyone who normally do the measurements – therefore, if there are several teams or shifts who do the measurements, try to ensure they are represented.
- It is also important to have an observer (or facilitator) – someone who won't take part in the actual study, but who will observe the study, coordinate the measurement of the parts and note the results. Their role will be discussed later in this section.
- If possible, arrange to have some help from a Black Belt when planning the study, and also when analysing and interpreting the results (especially the first few times you conduct a Gauge R&R).

It is recommended that when running an Attribute Agreement Analysis wherever possible to have an 'expert' assess each of the parts, in addition to the three appraisers we have already mentioned. Where it is possible, use an 'expert' (or a group of 'experts') to assess each part and determine the correct result for each. The expert decision is referred to as the **Standard.** This can then be used to assess the level of agreement between the appraisers and the agreed standard.

SEEK GUIDANCE

If unsure which 3 appraisers to choose, or if you feel that it might be better to have more (or less) than 3 people (e.g. if you have 4 teams doing the measurements), discuss with a local Black Belt.

If you are unsure how to determine the 'standard' result for attribute data, also discuss with a local Black Belt.

Continuous Data

Selecting Parts for the Study

With Alison the Black Belt's help, Julie and the team select 10 screens from those processed during the previous 24 hours – they select every 20th screen processed, as they know from their knowledge of the screen manufacturing process that this is likely to cover the full range of thicknesses typically produced, from 1.5mm to 2.0mm. They agree that Kevin, Mary and Karen will each measure each of the 10 screens 3 times.

The team are now looking forward to conducting the study and ready to begin.

Attribute Data

Selecting Parts for the Study

Three of the technicians who normally do the in-line inspection are chosen to take part in the study – Steven, Jim and Emma. These 3 technicians are chosen from a possible 8, as they represent all 3 shifts and are of varying levels of experience.

Again with the assistance of Alison the Black Belt Tom selects a sample of 20 screens from those recently processed. They ensure that a selection of good and bad screens are chosen, including a few which are clearly borderline (i.e. quite difficult to decide whether to pass or fail).

Although it isn't possible to label the screens, Tom feels that he will be able to keep track of them as an observer. He puts the screens into numbered boxes between measurements, and decides that he will randomly introduce them into the inspection queue when the study is being carried out. Alison agrees to assist with the observation in that she will monitor the database and extract the results as they are entered.

Tom and Alison decide that each technician will inspect each screen a total of 3 times.

In order to generate a standard set of results, they also decide that Mark, the final test engineer will assess each of the screens once – he is seen as the expert in determining whether a screen is actually a pass or a fail. Mark's assessment will be the standard against which the others will be compared to.

2c) Deciding when to run the study

We also need to consider when to run the study. In some cases, running the study will be quite straightforward, and it can simply be done at a time which suits the appraisers and the observer. However, it is important to consider the following when arranging a time:

- Will you be using any software, databases etc which might interfere with the normal running of the process – if so, could you run the study at the end of the day, or before the process starts in the morning?
- How long will it take to assess each part? Will you need to plan the study over a few days, or can it all be done at once?
- We will later discuss the need for the appraisers to measure the parts for a second and third time without being able to remember their previous results – this may mean that the study needs to be spread over a few days, rather than the 3 sets of measurements being done in a single day.

Deciding When to Run the Study

Julie and the team estimate that the study will take a maximum of 2 hours to run. They decide to do the first set of measurements in the morning, the second set immediately after lunch and the third set at the end of the day.

Continuous Data

2d) Planning the data collection

We can now start to plan the data collection:

- If possible, set up a Minitab worksheet, or excel spreadsheet, into which the observer can type the data directly to save time and possible errors.
- If this isn't possible, print out a blank Minitab worksheet for recording the data.
- The worksheets for Gauge R&R and Attribute Agreement Analysis each have their own layout.

SIGNPOST

For detailed guidance on how to set up and understand the layout of Minitab worksheets for Gauge R&R and Attribute Agreement Analysis go to ***Appendix 1***

2e) Random-isation of the study

When Minitab is used to set up the worksheet then it will automatically randomise the data. This means putting the measurements into a random order.

Randomising the parts before each of the sets of measurements is important, as this ensures that if there are any time-related or environmental causes of variation in the measurement system, these will be spread randomly throughout the study.

For example, some possible time-related changes could include people getting tired towards the end of the study or equipment heating up as the day progresses.

Where Minitab is not available to randomise the order of measurement then the random order can be determined simply by writing the numbers 1-10 or 1 – 20 on pieces of paper, and selecting them one at a time from a 'hat' or a box.

SIGNPOST

For detailed guidance on the options of randomising the data in Minitab worksheets for Gauge R&R and Attribute Agreement Analysis go to ***Appendix 1***

SEEK GUIDANCE

Help

If you are unsure about using Minitab or randomising the study then contact a local Black Belt for guidance.

Step 3

Conducting the Study

3a) Collecting the data
3b) Role of the observer
3c) Enter the data

3a) Collecting the data

Now we have the parts, the appraiser and the data collection sheet in place we are ready to conduct the MSA study.

Following the randomised order created on the worksheet each part should be measured /appraised three times by each of the appraisers.

It is critical that the result of each measurement is not influenced by previous measurements – the appraisers must **not** know the results for their previous measurements or for the measurements of the other appraisers otherwise they may accidentally bias the results (i.e. change their results so that they match the previous ones).

Ideally the appraisers should be unaware that they are even measuring the same parts or items more than once – this is often described as the measurements being done 'blind'. The observer has an important role to play in this, which we'll discuss on the next page.

Randomise the Study
Julie discusses how to randomise the study with Alison (the Black Belt). They decide to put each of the screens in a numbered box (numbered from 1-10). Alison will be the observer in the study and will hand the screens to the technicians for the measurement one at a time, to ensure that the technicians don't recognise them from previous measurements. They decide to use the Minitab worksheet to randomise the order of the trials.

Continuous Data

Randomise the Study
Tom and Alison also use Minitab to randomise the order of measurement of the screens, and are confident that measurements will be done 'blind' as the technicians are unlikely to realise that they are inspecting the same screen more than once.

Attribute Data

 |

3b) Role of the Observer

The observer's role in conducting the MSA is key. They will be required to:

- Plan, or be involved in planning the study.
- Identify which parts/items are which, in such a way that the appraisers aren't aware of any labelling or marking. This can be done in a number of ways:
  - If physical items are being assessed (e.g. manufactured parts, drawings or invoices), label each part discretely so that only the observer can see the label. Alternatively, for smaller items, place each item in a labelled box, and pass the items one-by-one to the appraisers without the label being seen.
  - For items such as recorded calls, save each call in a separate file or on a separate tape or CD, and again ensure that the appraisers can't see the label/filename.
- Ensure that the items are in random order before each set of measurements begin. Pass the items to the appraiser one-by-one for measurement
- Note the results, ensuring that none of the appraisers can see the results, or hear each other's results.
- Enter the results into the Minitab worksheet immediately if possible.
- Ensure that the appraisers are referred to as A, B and C, or 1, 2 and 3 – names should be noted (to allow for investigation of any unusual results) but not published.
- Note any comments during the study – anything of note which happens while the measurements are being made. For example:
  - If the measurement equipment or software has a fault or error during the measurement and needs to be corrected or replaced
  - If one of the parts or equipment is dropped
  - If the appraiser has any problems in making the measurement or if there are any noticeable differences in the way the appraisers make the measurements
  - If there are any interruptions or distractions
  - If there are any obvious opportunities for improvement.

3c) Entering the data

During or immediately after the study the data should be entered into Minitab.

Whilst it is most efficient for the data to be directly entered into Minitab it is important to be aware that it is quite easy to accidentally delete, mix up or mistype data. It is therefore advisable to write the measurements down prior to entering them into the Minitab worksheet

SIGNPOST

For detailed guidance on how to avoid the common pitfalls associated with entering the data into Minitab please go to ***Appendix 2***

Step 4

Interpreting the Results for Continuous Data

4a) Analysing the data in Minitab
4b) Interpretation of graphical output
4c) Interpretation of numerical output
4d) Communicating the results

Now the data has been collected we are ready to start to analyse, interpret and communicate the results

As the methods for analysis are very different for Gauge R&R and for Attribute Agreement Analysis we will take each in turn using Julie and Tom's case studies as examples.

We will begin with Gauge R&R (for continuous data) and then move on to Attribute Agreement Analysis (for attribute data). If you only wish to follow the procedure for Attribute Agreement Analysis then please turn to **page 48** now.

4a) Analysing the data in Minitab

Case Study – Data Analysis

Open the file Julies GAUGE R&R DATA 1. This is the data collected from Julie's measurement study.

To start with it is worthwhile just to look at the results. Can you see any difference in the results:

- With the same technician measuring the same screen repeatedly?
- With different appraisers measuring the same screen?

For example part 5 measured by appraiser 1 (Kevin) was measured at 1.55mm whereas this part measured by appraiser 2 (Mary) was measured at 1.41mm

↓	C1	C2	C3	C4
	RunOrder	Parts	Appraiser	Measurement
1	1	5	1	1.55
2	2	10	1	1.65
3	3	9	1	2.00
4	4	6	1	2.00
5	5	2	1	2.00
6	6	4	1	1.94
7	7	8	1	1.85
8	8	7	1	1.95
9	9	1	1	1.63
10	10	3	1	1.83
11	11	4	2	1.79
12	12	1	2	1.56
13	13	7	2	1.95
14	14	3	2	1.80
15	15	5	2	1.41
16	16	2	2	2.07
17	17	9	2	2.05
18	18	6	2	2.00
19	19	8	2	1.73
20	20	10	2	1.60
21	21	3	3	1.80
22	22	7	3	1.95
23	23	2	3	2.00
24	24	6	3	2.00

*To learn how to use the Minitab Gauge R&R function to analyse this data turn to **Appendix 3***

Minitab uses a technique called 'Analysis of Variance (ANOVA)' to analyse the data collected during the study.

If we have a measurement system which is performing well, and consistently, there will be very little difference between the repeated measurements made on each of the parts during the study – we describe this as there being very little **variation** in the measurement system.

If the measurement system is not performing well, there will be noticeable (or significant) variation – i.e. there will be a difference between the repeated measurements made on each of the parts.

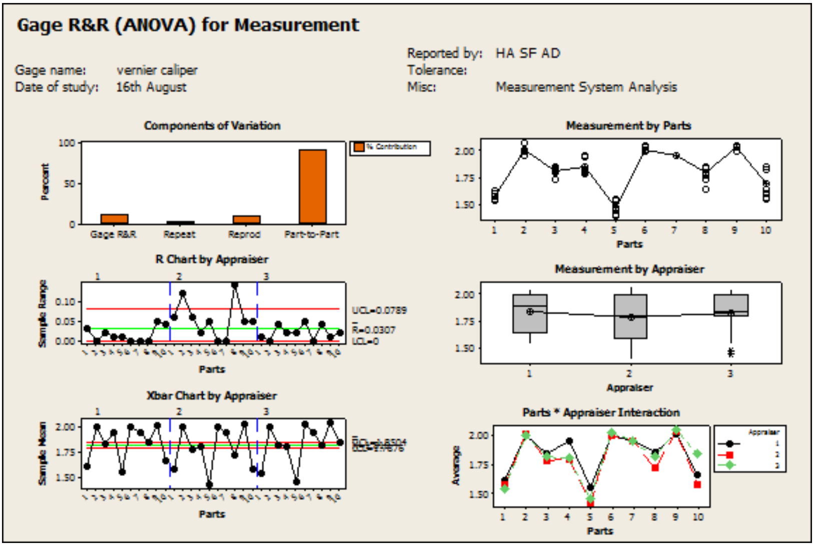
The ANOVA technique quantifies the amount of variation in the measurement system so that we can easily decide whether the measurement system is acceptable for use or not (using some guidelines which we will discuss shortly.)

Minitab also analyses the data in a such a way that, if there is an unacceptable amount of variation in the measurement system, we can easily see where that variation is coming from, by using the graphs which are produced, which will help us find the root cause(s) of the variation between measurements.

Minitab can either analyse the data and produce a numerical output or a set of graphs. Both are used in tandem to analyse the data. Each of these methods will be explained on the following pages.

4b) Interpret graphical output

First of all we will look at the graphs which are produced by Minitab. We will look at each of the graphs shown below in turn.



SIGNPOST

*For details on how to set up and produce the graphs in Minitab see **Appendix 3***

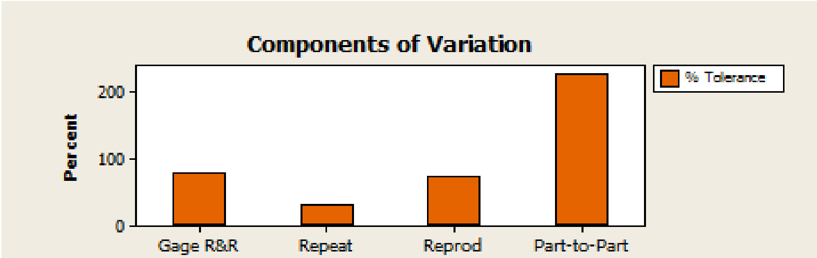
*Further information on interpreting each of the graphs is given in **Appendix 4***

SEEK GUIDANCE

Help

If you are unsure as to how to construct or interpret any of the graphs, ask for help from a local Black Belt.

The first graph to look at is the "**Components of Variation**" graph.



This is the most important of the graphs to look at as it tells us how well the measurement system is performing overall. For a good measurement system, we would expect the Gauge R&R and the Repeat and Reprod bars all to be very small compared to the Part-to-Part bar. This indicates that most of the variation seen in the study comes from genuine variation between the parts rather than from variation due to the repeatability or reproducibility of the measurement system. If there is a problem with the measurement system then the Gage R&R bar will be relatively tall representing 30% or more on the left hand scale.

In cases where the Gage R&R bar is too tall (as in the example above), we can also use this graph to help us determine whether the problems with the measurement system are due to Repeatability, Reproducibility or both. We look to see whether one of the bars labelled Repeat and Reprod is noticeably taller than the other, or whether they are similar in height.

- If the Repeat bar is taller, this indicates that there is an issue with Repeatability.
- If the Reprod bar is taller, this indicates that there is an issue with Reproducibility
- If they are a similar height, this indicates that there is an issue with both Repeatability and Reproducibility.

Interpreting Julie's Data – The Graphical Output

Julie and her team see that the Gauge R&R bar is quite tall compared to the Part-to-Part bar, which indicates that a significant proportion of the variation they are seeing in their measurements is coming from variation in the measurement system.

They also see that the Reprod bar is quite a bit taller than the Repeat one – this suggests that most of the causes of the measurement system being poor are likely to be related to Reproducibility.

They continue to look at the rest of the graphs to see if they can find some clues as to what the issues are.

WORKBOOK EXERCISE

Turn to **Exercise 6** in the Workbook to learn more about interpreting the components of variation graph

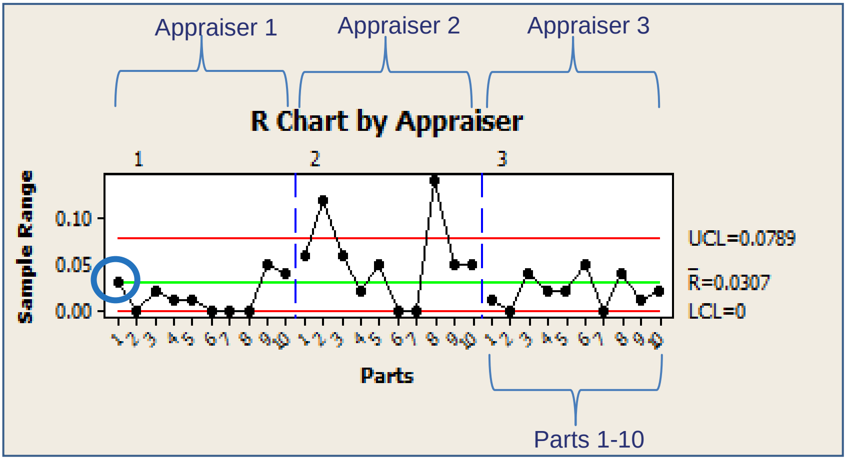
SIGNPOST

*For further information on interpreting this graph, refer to **Appendix 4***

Next we will look at the **R Chart by Appraiser**.

This chart shows the **R**ange of the results for each appraiser for each of the 10 parts.

There are 3 segments in the graph – one for each appraiser.



In each appraiser's segment, the **range** is plotted for each of the parts measured. (You will see the part numbers (1-10) along the bottom axis, repeated 3 times – once for each appraiser)

So in the above example, the first point plotted is the range of Kevin (appraiser 1)'s results for part number 1.

↓	C1	C2	C3	C4
	RunOrder	Parts	Appraiser	Measurement
1	9	1	1	1.63
2	32	1	1	1.60
3	69	1	1	1.60
4	12	1	2	1.56

The range is the maximum value minus the minimum value, so if we look at Kevin's results for part number 1 in the Minitab worksheet, we can see that his measurements were 1.63, 1.60 and 1.60.

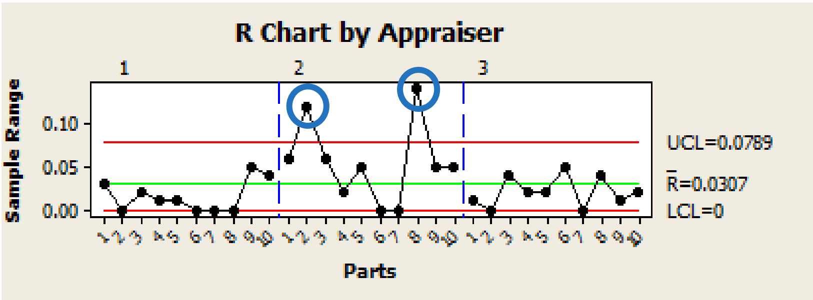
So the range was 1.63-1.60=0.03. We can see this plotted as the first point on the graph.

SIGNPOST

For further information on interpreting this graph, refer to ***Appendix 4***

Interpreting the R-Chart by Appraiser

We interpret this chart by saying that if any of ranges is particularly large (above the upper red line) then that indicates there is inconsistency in the measurements made by that particular appraiser on that particular part.



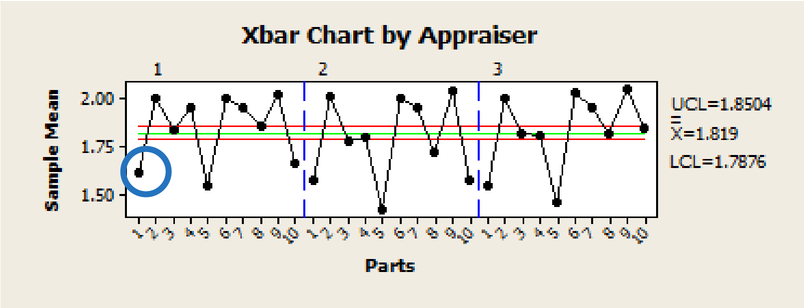
High ranges are often due to typing errors – check for this first.

Interpreting Julie's Data – The Graphical Output

On studying the R Chart by Appraiser from their study, Julie and her team can see that there are two points with high ranges (above the upper red line) - these correspond to screen numbers 2 and 8 when measured by Mary (appraiser 2). They check for typing errors, but are confident that the data has been entered correctly. They note this observation and move on to look at the next graph.

Continuous Data

Next we will look at the **Xbar Chart by Appraiser**.



As with the previous chart, there is a segment for each appraiser. In each appraiser's segment, the **average** measurement for each of the parts is plotted.

So in the above example, the first point plotted is the average of Kevin (appraiser 1)'s results for part number 1.

↓	C1	C2	C3	C4	C
	RunOrder	Parts	Appraiser	Measurement	C
1	9	1	1	1.63	
2	32	1	1	1.60	
3	69	1	1	1.60	
4	12	1	2	1.56	A

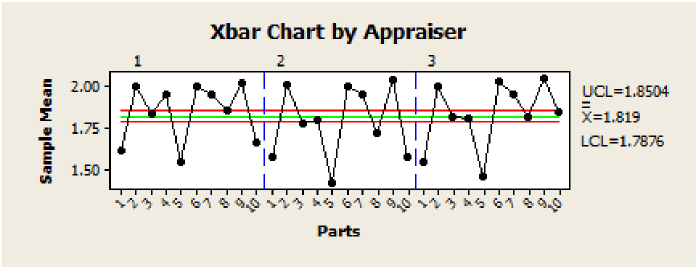
Looking at Kevin's results for part number 1, we can see that his measurements were 1.63, 1.60 and 1.60, therefore the average is 1.61. We can see this plotted as the first point on the graph.

SIGNPOST

*For further information on interpreting this graph, refer to **Appendix 4***

Interpreting the Xbar chart by Appraiser

If the measurement system is good, and the appraisers are consistent with each other, we would expect to see at least 50% of the points either above or below the red lines (and very few points between the red lines).

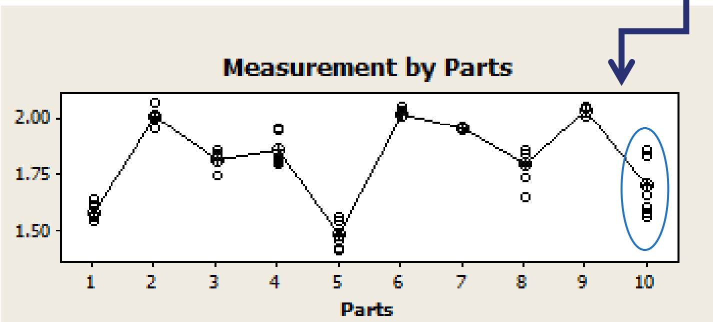


Julie notes that most of the parts are outside the control limits and so moves on to look at the next graph

Although this chart is an X-Bar Chart it is NOT being used in this context to assess control. The red lines in this case are NOT process control limits but rather represent the variation due to the measurement system. Therefore in this case it is desirable for the points to be outside the red lines.

Next we will look at the **Measurement by Parts** graph.

9 measurements
and average value
for part number 10



The scale of the Y-axis for this graph is not labelled in Minitab but represents the measurement (in this case the measurement in mm for each part)

This graph shows a circle for each measured value of each part (i.e. all measurements made by all appraisers). So in the case study example there are 9 circles for each part, as each part was measured 9 times in total. The average value for each of the parts is also shown by a crossed circle.

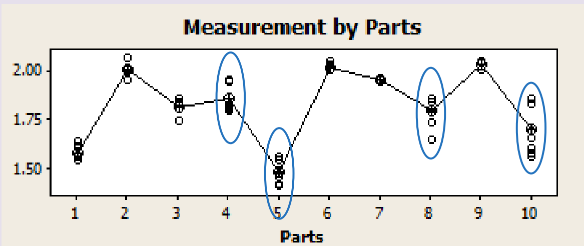
Interpreting the Measurement by Parts Graph

If the measurement system is consistent, there should be very little difference between the measurements for each individual part (in other words, the circles for each part should almost be on top of each other or overlapping).

If any of the parts has a noticeably larger spread in results than the others, this particular part might be worth investigating.

Similarly any parts with noticeably less spread in results than the others (such as part 7 below) may also be worth investigating as this can also indicate something unusual within the measurement system.

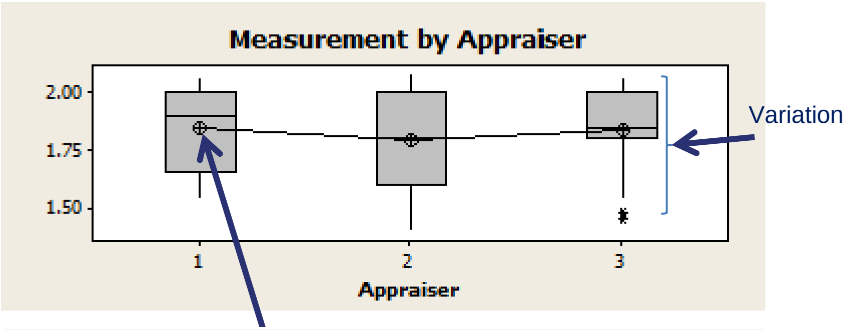
Julie and her team notice from this graph that screens 4, 5, 8 and 10 appear to have a greater spread than the rest of the screens. The team need to consider why these screens were more difficult to measure? They note this finding and move on to the next graph.



Continuous Data

*For further information on interpreting this graph, refer to **Appendix 4***

Next we will look at the **Measurement by Appraiser** graph.



Appraiser 1's measurements (Average= 1.84)					
1.63	1.60	1.60	2.00	2.00	2.00
1.83	1.85	1.83	1.94	1.95	1.95
1.55	1.55	1.54	2.00	2.00	2.00
1.95	1.95	1.95	1.85	1.85	1.85
2.00	2.00	2.05	1.65	1.65	1.69

In this graph, the crossed circles show the average of all of the results for all of the parts as measured by that appraiser. So for appraiser 1 (Kevin), the average of all of his measurements for all 9 screens was 1.84, as shown.

There is also a box plot for each appraiser – the represents the amount of variation in the results for all of the parts as measured by that appraiser. The middle 50% of the variation is represented by the height of the box and the full range of variation by the full length of the plot.

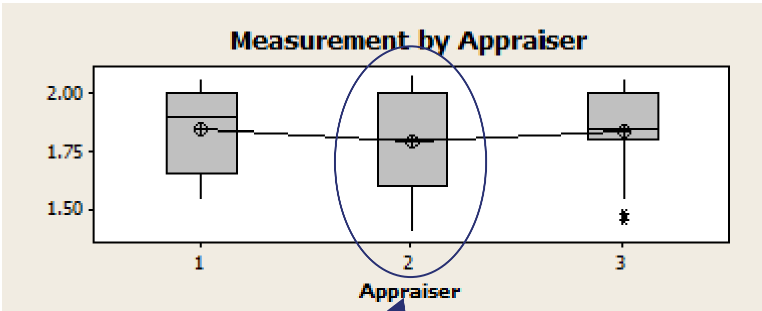
Interpreting the Measurement by Appraiser Graph

If the measurement system is consistent, we would expect the average values to be similar (as we would expect all appraisers to get very similar results for all of the parts). If this is the case, the line connecting the averages would be perfectly horizontal.

For the same reason, we would also expect the spread for each appraiser (represented by the size of the boxes and the length of the whiskers) to be similar.

If the line connecting the averages is not a straight line (i.e. if any of the appraisers has either a noticeably higher or lower average than the others), this suggest that they are perhaps doing something consistently different to the others (poor reproducibility)

If any of the appraisers has a noticeably larger variation than the others, this may indicate that that they doing something inconsistent from one measurement to the next (poor repeatability)

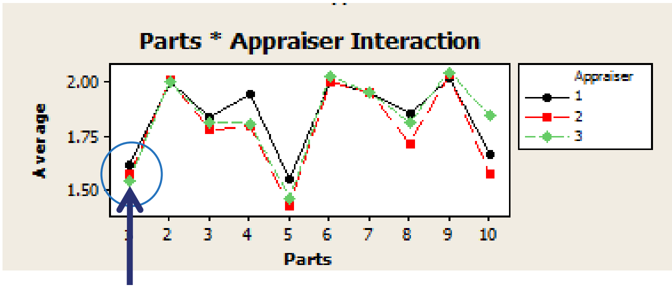


Julie and her team notice that appraiser 2 (Mary) has a slightly larger spread of results than Kevin and Karen, and also a lower average (the connecting line is not horizontal), which suggests that she might be doing something differently.

Again Julie and the team note their findings and move onto the final graph.

*For further information on interpreting this graph, refer to **Appendix 4***

Finally, we will look at the **Appraiser*Part Interaction** graph.



This graph overlays the average measurements for each item as measured by each appraiser. So for screen number 1, Kevin's average was 1.61 (dot), Mary's was 1.57 (square) and Karen's was 1.54 (diamond). The average measurements for each appraiser are connected by straight lines.

↓	C1	C2	C3	C4
	RunOrder	Parts	Appraiser	Measurement
1	9	1	1	1.63
2	12	1	2	1.56
3	26	1	3	1.54
4	32	1	1	1.60
5	43	1	2	1.55
6	55	1	3	1.54
7	69	1	1	1.60
8	72	1	2	1.61
9	82	1	3	1.55

Interpreting the graph

If the measurement system is perfectly consistent, we would expect all of the points for each part to be on top of each other (similar to part 2 in the example above), and the lines connecting would be parallel with each other.

We interpret this graph by saying that if the lines connecting the averages are not parallel, and there is a noticeable spread in the points for any of the parts, it is worth investigating why the measurements for these particular parts are less consistent than the others.

Julie and the team notice from this graph that the points are noticeably separate for screen numbers 4, 5, 8 and 10, and the connecting lines aren't parallel – this confirms what they saw in some of the previous graphs.

Continuous Data

*For further information on interpreting this graph, refer to **Appendix 4***

4c) Interpretation of numerical output

Interpreting the Numerical Data

In addition to the graphical output Minitab also provides us with a lot of numerical data. However the only output we are interested in for now is the Total Gage R&R % Tolerance (as highlighted on the session window below).

Session

Gage R&R

Source	VarComp
Total Gage R&R	0.0041159
Repeatability	0.0005744
Reproducibility	0.0035415
Appraiser	0.0005886
Appraiser*Parts	0.0029529
Part-To-Part	0.0353663
Total Variation	0.0394822

Process tolerance = 0.5

Source	StdDev (SD)	Study Var (6 * SD)	%Tolerance (SV/Toler)
Total Gage R&R	0.064155	0.38493	76.99
Repeatability	0.023968	0.14381	28.76
Reproducibility	0.059510	0.35706	71.41
Appraiser	0.024260	0.14556	29.11
Appraiser*Parts	0.054341	0.32604	65.21
Part-To-Part	0.188059	1.12836	225.67
Total Variation	0.198701	1.19221	238.44

Number of Distinct Categories = 4

In the above example from the case study, you will see that the Total Gage R&R %Tolerance is 76.99% – this tells us that the variation in the measurement system is 76.99% of the process tolerance (upper tolerance limit-lower tolerance limit). [The process tolerance for the screen is 2.0mm – 1.5mm = 0.5mm]

How do we decide whether this is acceptable or not?

The guideline for whether a measurement system is acceptable or not, is

% Tolerance*	System is…
<10%	Ideal
10 – 20%	Acceptable
>20%	Unsatisfactory

* This is a guideline criteria remember to check the latest version of the SABRe standard for the limits for your application. This standard also sets out the requirements for bias which are outside the scope of a gauge R&R study.

SEEK GUIDANCE

If the %Tolerance is outside the acceptance criteria, the results should be discussed with a local Black Belt, who will help you assess the risks of continuing to use the system – remember, making decisions with an unreliable measurement system can cost your organisation a lot of money.

The other numerical output which we are interested in is the **Number of Distinct Categories.**

```
Number of Distinct Categories = 4
```

An acceptable measurement system will have 5 or more distinct categories. This means that it is able to discriminate between different parts – 5 distinct categories means that it can discriminate between 5 different sizes of parts across the total study variation

If the discrimination is poor, the gauge is poor at distinguishing parts from each other and will only classify them as, for example, high and low (2 distinct categories) , or high/medium/low (3 distinct categories). This is not adequate for process control and analysis.

Interpreting Julie's Numerical Data

So, as the % tolerance of 76.99% is greater than 30% and the number of distinct categories is <5, Julie and her team conclude that the measurement system is UNACCEPTABLE and in need of improvement.

Julie immediately contacts her local Black Belt Alison for advice on how to proceed.

4d) Communicating results

The results of the Gauge R&R can now be communicated as necessary.

Try and keep the communication simple –include the %Tolerance results, along with an explanation that >30% is deemed unacceptable, and that the results will need to be discussed with a Black Belt.

Also include one or two of the graphs in the communication – don't include too many though, choose the ones which clearly show what the problem is, and the ones which you find easiest to explain (as you will have noticed, in many cases the same problem will be highlighted by several of the graphs)

Summary – Interpretation of Minitab Output

Julie and the team get together to consider the results of the study and to discuss possible opportunities for improvement

Their summary of the study is:

- The %Tolerance of 76.99% is unacceptable and the measurement system should not be used until it has been improved.
- Poor Reproducibility indicates that the technicians are not consistent with each other in their measurement of screen thickness.
- Mary was the least Repeatable in measuring the screen thickness. The Range Chart indicates that Mary was particularly inconsistent in her measures for parts 2 and 8.
- There appears to be a particular problem with screen numbers 4,5, 8 and 10

Continuous Data

Once the results of the Gauge R&R Study have been interpreted action must be taken to address any repeatability and/or reproducibility issues identified by the study as unacceptable.

*To continue to follow the continuous data Gauge R&R case study please turn to **page 61***

Step 4

Interpreting the Results for Attribute Agreement Analysis

4a) Analysing the data in Minitab
4b) Interpretation of graphical output
4c) Interpretation of numerical output
4d) Communicating the results

We will now examine how to analyse, interpret and communicate the results for **Attribute Agreement Analysis**

4a) Analysing the data in Minitab

Attribute Data

Tom and Alison carry out the study, and Tom notes some interesting observations. Tom enters the data into Minitab and is now ready to analyse the results. Open Minitab file **TOMS ATTRIBUTE DATA 1** to view the results worksheet which includes Tom's observations.

We can now begin to analyse the data in Minitab.

> *For detailed guidance on how to carry out an Attribute Agreement Analysis in Minitab go to **Appendix 8***

How does Minitab analyse the data?

If we have an attribute measurement system which is working perfectly, each appraiser will get the same result (or come to the same decision) for every part every time they assess it:

- Each appraiser's results will be in 100% agreement (this is conceptually the equivalent of Repeatability).
- All of the appraisers will get the same results or reach the same decisions as each other – they will be in 100% agreement with each other (this is conceptually the equivalent of Reproducibility).
- Their results will all be in 100% agreement with the 'expert' or 'standard' (this is the equivalent of Bias)

Minitab analyses the results of the Attribute Agreement Analysis and tells us what percentage of the time the appraisers agree with themselves and also what percentage of the time they agree with each other – these 2 percentages allow us to assess both the **% With Appraiser Agreement** and **% Between Appraiser Agreement** for the measurement system.

It also tells us what percentage of the time the appraisers are in agreement with the Standard.

Having determined the percentages described above, we then use some guidelines to determine whether the measurement system needs to be improved or not.

Below is an extract from Tom's worksheet –
Column 2 contains the number which was designated to each screen.
Column 3 contains the number of the technician who inspected the screen:

- Appraiser 1 is Steven
- Appraiser 2 is Jim
- Appraiser 3 is Emma

Column 4 contains the decision (pass/fail) made by the technician assessing each screen.
Column 5 contains the decision made by Mark (final test engineer).

Attribute Data

↓	C1	C2-T	C3-T	C4-T	C5-T
	RunOrder	Screen Number	Technician	Measurement	Standard
1	1	15	1	Fail	Fail
2	2	4	1	Pass	Pass
3	3	16	1	Pass	Fail
4	4	10	1	Fail	Fail
5	5	19	1	Pass	Pass
6	6	14	1	Pass	Pass
7	7	11	1	Pass	Fail
8	8	5	1	Pass	Pass
9	9	9	1	Pass	Pass
10	10	7	1	Pass	Fail
11	11	8	1	Fail	Fail
12	12	12	1	Pass	Pass
13	13	1	1	Pass	Fail
14	14	18	1	Fail	Fail
15	15	3	1	Pass	Fail
16	16	2	1	Pass	Pass
17	17	13	1	Pass	Pass
18	18	20	1	Fail	Fail
19	19	6	1	Pass	Pass
20	20	17	1	Pass	Pass
21	21	6	2	Pass	Pass
22	22	8	2	Fail	Fail
23	23	11	2	Pass	Fail
24	24	10	2	Fail	Fail
25	25	4	2	Pass	Pass
26	26	14	2	Pass	Pass

SIGNPOST

*Refer to **Appendix 8** for a guide to analysing the data in Minitab*

Before analysing the data in Minitab, Tom and Alison have a look at the results. They immediately notice some inconsistencies, a few of these can be seen in the worksheet below are:

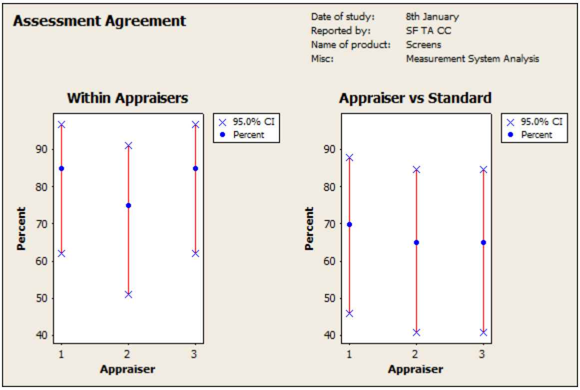
In row 3, screen 16 was assessed by Mark as a Fail, but Steven classified it as a Pass.

In row 10, screen 7 was also assessed as Fail by Mark and a Pass by Steven.

↓	C1	C2-T	C3-T	C4-T	C5-T
	RunOrder	Screen Number	Technician	Measurement	Standard
1	1	15	1	Fail	Fail
2	2	4	1	Pass	Pass
3	3	16	1	Pass	Fail
4	4	10	1	Fail	Fail
5	5	19	1	Pass	Pass
6	6	14	1	Pass	Pass
7	7	11	1	Pass	Fail
8	8	5	1	Pass	Pass
9	9	9	1	Pass	Pass
10	10	7	1	Pass	Fail
11	11	8	1	Fail	Fail
12	12	12	1	Pass	Pass
13	13	1	1	Pass	Fail
14	14	18	1	Fail	Fail
15	15	3	1	Pass	Fail
16	16	2	1	Pass	Pass
17	17	13	1	Pass	Pass
18	18	20	1	Fail	Fail
19	19	6	1	Pass	Pass
20	20	17	1	Pass	Pass
21	21	6	2	Pass	Pass
22	22	8	2	Fail	Fail
23	23	11	2	Pass	Fail
24	24	10	2	Fail	Fail
25	25	4	2	Pass	Pass
26	26	14	2	Pass	Pass

4b) Interpretation of graphical output

As with Gauge R&R, Minitab produces both numerical and graphical output (turn to Appendix 3 for a guide to how to produce these) – we will look at the graphical output first.

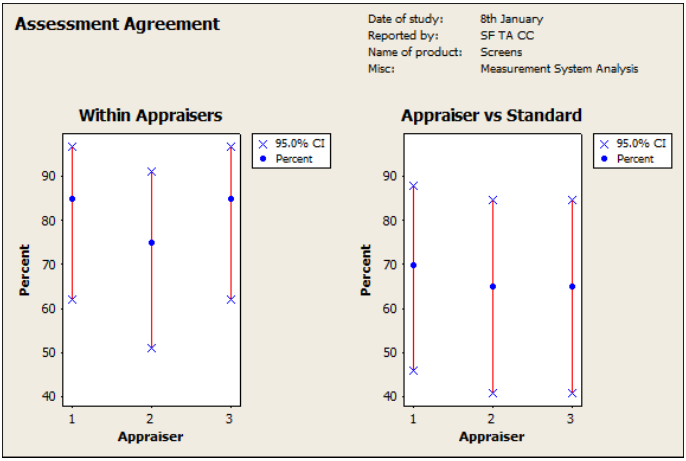


The left hand graph (Within Appraisers) shows the agreement of the appraisers as a percentage score. The dot for each appraiser shows the percentage of the measurements where the appraiser agrees with him or herself.

So, in the case study example above, appraiser 1 (Steven) agrees with himself 85% of the time – this means that of the 20 parts inspected, he made the same decision every time for 17 of the parts (85% of 20 = 17). However his decisions weren't consistent for the other 3 parts (if you look at the worksheet you will see that his decisions were inconsistent for parts 3, 4 and 16).

	RunOrder	Screen Number	Technician	Measurement	Stan
1	15	3	1	Pass	Fail
2	80	3	1	Fail	Fail
3	138	3	1	Fail	Fail
4	2	4	1	Pass	Pass
5	67	4	1	Pass	Pass
6	130	4	1	Fail	Pass
7	3	16	1	Pass	Fail
8	72	16	1	Fail	Fail
9	140	16	1	Fail	Fail
10					

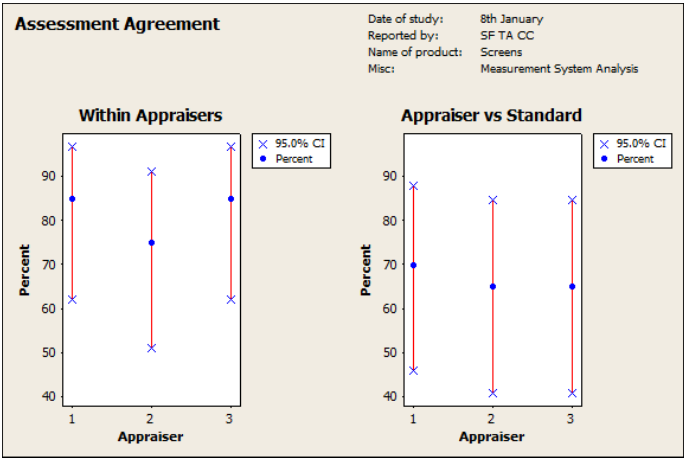
*Refer to **Appendix 8** for information on interpreting the crosses and red lines shown on the graph above – these represent confidence intervals.*



The right hand graph (Appraiser vs. Standard) shows how well each of the appraisers agrees with the standard (i.e. in the case study, how often each of the appraisers reached the same decision as Mark, who is seen as the expert). In the example above, we can see that appraiser 2 (Jim) agreed with Mark 65% of the time – this means that of the 20 parts measured, Jim reached the same decision as Mark on 13 occasions (65% of 20=13). If you look at the worksheet you will see that on screens 1, 3, 4, 7, 9, 11 and 16, Jim's decision disagreed with Mark's on at least one of the 3 inspections.

Final(Technician = 2) ***

↓	C2-T	C3-T	C4-T	C5-T
	Screen Number	Technician	Measurement	Standard
1	1	2	Pass	Fail
2	1	2	Pass	Fail
3	1	2	Pass	Fail
4	3	2	Pass	Fail
5	3	2	Pass	Fail
6	3	2	Fail	Fail
7	4	2	Pass	Pass
8	4	2	Pass	Pass
9	4	2	Fail	Pass
10	7	2	Fail	Fail
11	7	2	Fail	Fail
12	7	2	Pass	Fail
13	9	2	Pass	Pass
14	9	2	Pass	Pass
15	9	2	Fail	Pass
16	11	2	Pass	Fail
17	11	2	Pass	Fail
18	11	2	Pass	Fail
19	16	2	Pass	Fail
20	16	2	Fail	Fail
21	16	2	Fail	Fail



Interpreting the Results

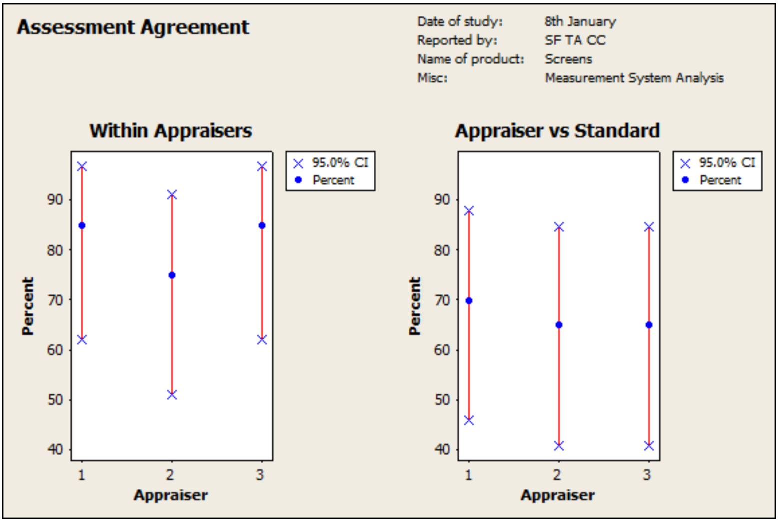
Ideally we would like 100% agreement in all cases (both ‘Within Appraisers’ and ‘Appraisers vs. Standard’); however, this is unlikely to be the case, so we use the following rules of thumb:

Percentage	Guideline
Greater than 90% on average	Excellent
70% - 90% on average	Acceptable
Less than 70% on average	Unacceptable

An unacceptable ‘Within Appraiser’ score means that the appraiser is not making consistent decisions with him/herself when repeatedly measuring some of the parts.

An unacceptable ‘Appraiser vs. Standard’ scores means that the appraisers are not making the same decision as the expert on some of the parts.

When using the above general guidelines you must take into account the risk associated with the decision being made by the measurement system. For example whilst 80% agreement may be acceptable for the assessment of expenses claims it may be unacceptable for the pass/fail decision on a critical part



Tom and Alan can see from these graphs that the % Within Appraiser Agreement (equivalent of Repeatability) for appraisers 1 and 3 is 85% and appraiser 2 is 75%, all of which are above the 'acceptable' guideline.
The 'Appraiser vs. Standard' results for Steven (70%) is acceptable; however Jim and Emma's results (both 65%) are not acceptable – their decisions are not consistent with Mark's, therefore the root cause of this needs to be investigated.

They can see some of the inconsistencies in the worksheet:

Attribute Data

nal(Technician = 3) ***

C2-T Screen Number	C3-T Technician	C4-T Measurement	C5-T Standard
1	3	Pass	Fail
1	3	Pass	Fail
1	3	Pass	Fail
3	3	Pass	Fail
3	3	Pass	Fail
3	3	Pass	Fail
11	3	Pass	Fail
11	3	Pass	Fail
11	3	Pass	Fail
12	3	Fail	Pass
12	3	Fail	Pass
12	3	Fail	Pass

These are some of the results obtained by Emma (technician 3). We can see several instances where she consistently reaches a different decision from the standard (Mark) – it looks like she particularly has a problem with screens 1, 3, 11 and 12.

4c) Interpretation of numerical output

Minitab also produces some numerical output in the session window – this provides more detail on the results seen in the graphs.

The first part of the output we will look at is the 'Within Appraisers' assessment agreement.

This gives us the % of times that each appraiser agreed with themselves consistently on their judgement about the part.

Within Appraisers

Assessment Agreement

Appraiser	# Inspected	# Matched	Percent	95% CI
1	20	17	85.00	(62.11, 96.79)
2	20	15	75.00	(50.90, 91.34)
3	20	17	85.00	(62.11, 96.79)

Matched: Appraiser agrees with him/herself across trials.

These are the percentages which we saw on the 'Within Appraisers' graph.

Next we look at how well the appraisers agreed consistently with the standard. This is called "Each Appraiser vs. Standard""

Each Appraiser vs Standard

Assessment Agreement

Appraiser	# Inspected	# Matched	Percent	95% CI
1	20	14	70.00	(45.72, 88.11)
2	20	13	65.00	(40.78, 84.61)
3	20	13	65.00	(40.78, 84.61)

Matched: Appraiser's assessment across trials agrees with the known standard.

These are the percentages which we saw on the 'Appraiser vs. Standard' graph.

The final piece of output we are interested in here is the Between Appraisers assessment agreement. This gives us the agreement expressed as a percentage of how consistently the appraisers' decisions agreed with each other.

Between Appraisers

```
Assessment Agreement

# Inspected  # Matched  Percent       95% CI
       20          13    65.00  (40.78, 84.61)

# Matched: All appraisers' assessments agree with each other.


Fleiss' Kappa Statistics

Response     Kappa   SE Kappa        Z  P(vs > 0)
Fail      0.679038  0.0372678  18.2205     0.0000
Pass      0.679038  0.0372678  18.2205     0.0000
```

Again, we look at the percentage. This percentage tells us for how many parts the appraisers agree with each other's decisions. In this example, they agree for 65% of the parts – so of the 20 parts inspected, they all consistently made the same decision on 13 of those parts. For the remaining parts, their decisions differed in some way (more detailed information on this can be obtained by studying the worksheet).

The rule of thumb for interpreting the percentages is as before:

Percentage	Guideline
Greater than 90% on average	Excellent
70% - 90% on average	Acceptable
Less than 70% on average	Unacceptable

An unacceptable 'Between Appraiser' scores means that the appraisers are not making consistent measurements with each other when repeatedly measuring some of the parts (the equivalent of poor reproducibility)

Tom and Alan can see that the percentage agreement between appraisers is 65%, so they conclude that the Measurement System is Unacceptable – the 3 technicians are Inconsistent with each other in making decisions on whether to pass or fail the screens.

Attribute Data

Fleiss Kappa Statistic

The Minitab session window also shows us some statistics known as **Fleiss Kappa statistics** for each of ‘Within Appraisers’, ‘Appraiser vs Standard’ and ‘Between Appraisers’. The statistic we are interested in is ‘Kappa’, as highlighted below. These statistics gives us an alternative measure for assessing how well the appraisers agree with themselves, the standard and each other. Rather than simply looking at the observed agreement, the Kappa statistic calculates the likelihood that the agreements seen could have occurred by chance.

- The Kappa Statistic estimates the probability that agreements occur by chance.
- The Kappa value ranges from -1 to 1
- The higher the Kappa value, the stronger the agreement
  - Kappa = 1 means perfect agreement
  - Kappa = 0 means the same as would be expected by chance (50:50)

The guidelines for interpreting Fleiss’ Kappa Statistics are as follows:

> 0.75	Excellent
0.4 – 0.75	Fair to Good
< 0.4	Poor Agreement

- So if we look at the Kappa values for the ‘Within Appraisers’ (from the case study), we see that appraisers 1 and 3 have Kappa values of >0.75 which is excellent, whereas appraiser 2 (Jim) has Kappa values of 0.641148, which is fair to good.

Within Appraisers

```
Assessment Agreement

Appraiser  # Inspected  # Matched  Percent      95% CI
1                   20         17    85.00  (62.11, 96.79)
2                   20         15    75.00  (50.90, 91.34)
3                   20         17    85.00  (62.11, 96.79)

# Matched: Appraiser agrees with him/herself across trials.


Fleiss' Kappa Statistics

Appraiser  Response     Kappa  SE Kappa        Z  P(vs > 0)
1          Fail      0.775000  0.129099  6.00312     0.0000
           Pass      0.775000  0.129099  6.00312     0.0000
2          Fail      0.641148  0.129099  4.96631     0.0000
           Pass      0.641148  0.129099  4.96631     0.0000
3          Fail      0.794286  0.129099  6.15251     0.0000
           Pass      0.794286  0.129099  6.15251     0.0000
```

Fleiss Kappa Statistic

Between Appraisers

Assessment Agreement

# Inspected	# Matched	Percent	95% CI
20	13	65.00	(40.78, 84.61)

Matched: All appraisers' assessments agree with each other.

Fleiss' Kappa Statistics

Response	Kappa	SE Kappa	Z	P(vs > 0)
Fail	0.679038	0.0372678	18.2205	0.0000
Pass	0.679038	0.0372678	18.2205	0.0000

- The Kappa values for the 'Between Appraisers', has Kappa values of 0.67 which again is fair to good.

Each Appraiser vs Standard

Fleiss' Kappa Statistics

Appraiser	Response	Kappa	SE Kappa	Z	P(vs > 0)
1	Fail	0.585240	0.129099	4.53325	0.0000
	Pass	0.585240	0.129099	4.53325	0.0000
2	Fail	0.591725	0.129099	4.58348	0.0000
	Pass	0.591725	0.129099	4.58348	0.0000
3	Fail	0.496215	0.129099	3.84367	0.0001
	Pass	0.496215	0.129099	3.84367	0.0001

- The Kappa values for 'Each Appraiser vs. Standard', has Kappa values as low as 0.49 for appraiser 3 which sits just within the fair to good guidance but indicates nevertheless the opportunity for improvement to the measurement system.

SIGNPOST

There are other variations, for example, for categorical data (see a Black Belt)

Summary – Interpretation of Minitab Output

- If all of the %Agreements are greater than or equal to 70%, the measurement system is acceptable, and can be used to collect data

- If any of the %Agreements are <70%, the graphs and further numerical data can be studied to help further understand and investigate the reasons for the variation in the measurements

- If any of the %Agreements are <70%, the measurement system is unacceptable, and should be discussed with a Black Belt

4d) Communicate results

The results of the Attribute Agreement Analysis can now be communicated as necessary.

Try and keep the communication simple – include the %Agreement results, along with an explanation that <70% is deemed unacceptable, and that the measurement system is not fit for purpose until improvements have been made.

Also include the graphs in the communication to help explain how the problems were found and also to help demonstrate the improvements.

Summary of Minitab Output

Tom talks through the results with Alison, Alan and the 3 technicians who were involved in the study. In his communication he thinks carefully about how to explain the results of the analysis to team. He stresses that the problems highlighted by the study are a reflection of weaknesses in the systems, standards, procedures and training being used rather than in the people themselves. He also continues to keep the results of the study anonymous (referring to the appraisers only as 1, 2 and 3) so that the team focussed on the overall results rather than identifying themselves within the trial.

- The minimum result for Within Appraisers was 75%. This is above the 70% required, which means that the % agreement within appraisers is acceptable. This means that each of the technicians is fairly consistent with themselves.

- The % agreement between appraisers however is unacceptable, as the Between Appraisers result was 65%. This means that the technicians are inconsistent with each other, which suggests that the method (or the equipment) being used by the technicians to determine whether each screen is a pass or a fail differs between technicians.

- Appraiser 2 & 3's Appraiser vs. Standard scores were 'unacceptable' at 65%. This means that their measurements are not consistent with Mark's, which might suggest that the criteria for determining whether each screen is a pass or a fail is unclear.

- The above inconsistencies (or variation) in the measurement system are resulting in some scratched screens being passed further through the process, resulting in wasted processing costs and some good screens are being rejected resulting in unnecessary scrap.

Step 5

Taking Action if the Results are Unacceptable

5a) Results are unacceptable

In the case of both Gauge R&R and Attribute Agreement Analysis if the results are unacceptable, as previously discussed, the first thing to do is to discuss the results with a local Black Belt.

We can then use the findings from the Minitab graphs to help find the root causes of the variation in the measurement system.

If the causes aren't immediately obvious, techniques such as brainstorming and cause and effect can be used to list all of the potential causes, and then each of these can be investigated until the root causes are found.

5b) Common reasons for unacceptable measurement systems

Here are some common reasons for unacceptable measurement systems, and some questions which can be asked to help find the root causes:

- If there is an issue with a particular part, compare this part with the other parts to see if this leads to any possible causes of the variation
- If there is an issue with the measurement system as used by one or more of the appraisers, observe the appraisers and ask them for possible reasons.
- Are the operational definitions clear and being followed by everyone? Has everyone been trained how to use them?
- Are the results being recorded consistently – look for rounding errors such as some people rounding up, others rounding down. Is everyone recording to the same number of decimal places?
- Is everyone using the same version of the current standard? The same applies to checklists, drop down menus, other reference documents.
- Is one or more of the appraisers becoming tired, rushed, bored?
- Do the appraisers understand why they are doing the measurement? If not, they may not realise the importance of following the operational definition consistently.

- Are the appraisers comfortable when making the measurement? Are they having to stretch or twist to read the measurements or note the results?
- Is their eyesight OK?
- Are the lighting conditions adequate?
- If using software, is everyone using the same version and selecting the same programme?
- Do the appraisers load the parts consistently?
- Is the procedure for loading and setting up the parts very complex? Could it be made easier e.g. by using a jig?
- Are temperature and humidity levels stable?
- Does the equipment need to be left to stabilise after switching on?
- Could the gauge or equipment be worn? Corroded? Contaminated?
- Could the method of conducting the Measurement System have had an impact on its results – for example could inspectors be looking extra hard for defects or did they spend much more time than normal measuring or inspecting the parts?

Use the observations noted during the test to look for any clues.

WORKBOOK EXERCISE

Turn to **Exercise 7** in the Workbook to identify further possible reasons for your own measurement systems

After further discussion and observation of the inspections done both at the in-line inspection and at final test, Tom and his team feel that the following were the root causes of the problems seen –

- One of the inspection lamps wasn't working properly, resulting in the light being too dim, and it was difficult to see some of the scratches. This was causing scratches to be missed, and screens with scratches being passed instead of scrapped.
- One of the sets of tweezers which was being used to hold the screens appeared to be scratching the screens – this was resulting in screens which passed at in-line inspection to fail at Final test.
- There seemed to be some confusion as to what a 'scratch' actually meant – some of the technicians thought that small scuffs were OK to pass, whereas others classed every mark on the screen as a fail.

Note: The team realised that they should have noticed these issues during their initial observations when preparing for the study (step 1). This highlighted to them how important the initial observation step is.

Attribute Data

The technicians and Alison discuss their observations from the results and the study, and feel that the following are possible root causes of the poor reproducibility and repeatability –

- The screens are quite small and difficult to hold. Each technician had a slightly different technique for holding them.
- Mary is left handed and she seems to find the vernier calliper more difficult to use because of this.
- One of the engineers took a closer look at screen numbers 4, 5, 8 and 10 and found that these screens are actually slightly warped, therefore the thickness result obtained depends on exactly where on the screen the measurement is taken.

Continuous Data

5c) Identify improvement opportunities

Identifying Improvement Opportunities

Once the causes of the variation have been found, we need to identify solutions which will remove these causes. Sometimes the solutions will be obvious, but sometimes solutions may need to be more creative. The ideal solution is one which makes the variation impossible – for example, the use of a jig to prevent parts from moving, or the use of software to check for errors.

Continuous Data

Identifying Improvement Opportunities

Julie and the team use brainstorming to identify some possible solutions to the problems identified.

- **Solution:** They decide to manufacture a jig in which to place the screens. The jig ensures that the screens are securely held, and also has markings to ensure that the thickness is measured in the same place on every screen (even if the screen is slightly warped). The jig is easy to use for both left and right-handed people.
- They also pass the warped screens to the Screen Manufacturing engineer for further investigation into the reasons for warping.

Attribute Data

Identifying Improvement Opportunities

Tom and his team use brainstorming to identify some possible solutions to the problems identified.

Solutions:

- New bright lights are purchased for the in-line inspection area, and a weekly check introduced to ensure that they are functioning correctly.
- The damaged tweezers are replaced as a short term solution – however a team is put together to look at developing a more suitable way of holding the screens without scratching them.
- Laminated photographs showing examples of passing and failing screens are placed next to the in-line inspection station, and the operational definition updated to reference these.

5d) Confirm the improvement

Confirming the Improvement

Once the improvements to the measurement system have been implemented, the MSA must be repeated (following the same procedure as before) to ensure that there is in fact an improvement to the results, and also to look for further improvement opportunities.

Continuous Data

Confirming the Improvement

Once the jig was manufactured, the team wrote a new operational definition for the measurement system and retrained all the technicians in its use.

In order to check whether the measurement system had in fact improved, they repeated the Gauge R&R study, using the same procedure as before.

Case Study – Data Analysis

Open the file **Julies GAUGE R&R DATA 2**. This file contains the data collected from Julie's new measurement study.

WORKBOOK EXERCISE

Re-run the analysis in Minitab and turn to **Exercise 8** in the Workbook to check if you have interpreted the Minitab output correctly

Attribute Data

Confirming the Improvement

Tom and his team train everyone involved in the use of the new operational definition, and ensure that the maintenance team have been trained in doing the weekly checks on the bright lights.

They then repeat the attribute agreement analysis, using the same sample of parts (where practical) and method as before.

Case Study – Data Analysis

Open the file **TOMS ATTRIBUTE DATA 2.** This file contains the data collected from Tom's new measurement study.

WORKBOOK EXERCISE

Re-run the analysis in Minitab and turn to **Exercise 9** in the Workbook to check if you have interpreted the Minitab output correctly

Step 6

Maintaining the Improvement

6a) Ensuring the gains are sustained

It is also important to ensure that the improvements are sustained, and that the performance of the measurement system is checked regularly.

6a) Ensuring the gains are sustained

Checking the Stability of the Measurement System

To monitor the ongoing stability of the measurement system, a regular (for example daily or weekly) measurement can be made of the same part and the results plotted on an SPC chart. This will detect any changes in the performance of the measurement system over time.

For more information on SPC Charts see the 'How To Guide' for SPC Charts.

The part which will be measured regularly is often referred to as the 'Reference Part', and should be stored carefully between measurements. The best reference part for this purpose is one that is 'borderline' as this will most effectively test the ability over time of the Measurement System to be able to discriminate between acceptable and unacceptable parts. It is recommended that a part is selected which falls near the middle of the tolerance limits, or it may be desirable to measure several parts which have low, high and mid-tolerance values. Alternatively, some measurement equipment is supplied with standard parts, which can also be used for this purpose.

Note: Measuring the same part on a daily basis may not be possible, or may not make sense, especially where attribute data is being collected. Asking someone to inspect the same item every day and determine whether it is a Pass or a Fail probably won't give meaningful results, as the appraisers will be biased by their previous results. Instead, it may be more useful to regularly choose a part at random from the process, ask both an appraiser and an 'expert' to assess it, and then compare their results, If their results are found to be different on any occasion, this indicates a problem with the measurement system, which should be investigated.

In either of the above cases, if a problem is found with the measurement system, it is recommended that the appropriate type of MSA is carried out as soon as possible, to further understand where the problem lies.

It is recommended that you discuss the most suitable frequency for ongoing stability checks with a local Black Belt

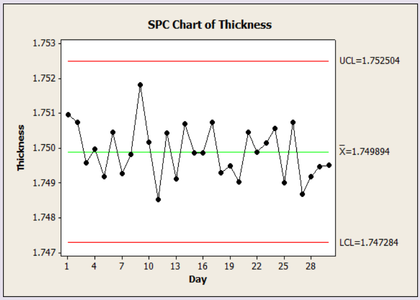
Checking the Stability of the Measurement System

Having confirmed the improvements, Julie and her team now need to think about how they will ensure that the improvements are sustained.

They select a screen which has a thickness very close to the target of 1.75mm, and decide that they will measure this screen on a weekly basis, to monitor the stability of the improved measurement system. They update the SOP for the process to state that this screen (which is referred to as the ‘Reference Screen’) will be measured every Monday morning, by one of the dayshift technicians, and the results plotted on an SPC Chart.

The data which they collect for the first 10 days is shown below, along with an SPC chart showing the results for the first month.

Day	Thickness
1	1.75098
2	1.75075
3	1.74958
4	1.74997
5	1.74919
6	1.75045
7	1.74928
8	1.74981
9	1.75182
10	1.75017



As you can see from the table, the measured thickness of the ‘Golden Screen’ for the first 10 days ranges from 1.74958mm to 1.75098mm, so all of the results are very close to the expected value of 1.75mm, which indicates that the measurement system is stable. This is also confirmed by the SPC chart, as all of the points are between the control limits (red lines).

Checking the Stability of the Measurement System

Tom and his team feel that inspecting a ‘Reference Screen’ for scratches on a daily basis probably won’t give meaningful results, as it will be very difficult to inspect the same screen repeatedly without being influenced by previous results. They did consider randomly introducing the ‘Reference Part’ into the inspection queue without the technicians being aware (as they did for the Gauge R&R); however, they are concerned that this will be difficult and time-consuming to implement and monitor. They are also concerned that it will be difficult to store the screen for a long period of time without it being scratched further.

Instead, they agree with Mark that, every Wednesday, he will take a screen at random from the production line, and assess whether he thinks it is a pass or a fail. He will then enter this result into a database, which will automatically compare his result to that obtained at the in-line inspection. The IT department are able to set up the database so that if the 2 results differ, a warning message will appear as soon as Mark enters his result, which will enable him to put the screen aside for further investigation. Tom will also automatically receive an email alerting him to the result.

Technician	Mark	Comparison
FAIL	FAIL	OK
PASS	PASS	OK
PASS	PASS	OK
FAIL	FAIL	OK
FAIL	FAIL	OK
PASS	FAIL	INVESTIGATE!
PASS	PASS	OK
PASS	PASS	OK
PASS	PASS	OK
PASS	PASS	OK
PASS	PASS	OK

Here are the results for the first 11 weeks. As you can see, on the 6th week, Mark and the technician don’t agree in their result. On investigation, it is quickly discovered that the technician who had done this inspection is a new employee, and they realise that, although they trained all of the existing quality technicians in the use of the new operational definition, they omitted to train one of the trainers, who trains new employees, This is quickly rectified – they re-train both the new employee and the trainer - and the training documentation is updated to include the new operational definition.

They then run an Attribute Agreement Analysis study, which confirms that the performance of the measurement system is acceptable, and continue with their weekly checks.

Repeating the MSA

For both continuous and attribute measurement systems, the MSA should be repeated at regular intervals (such as every 6 months or annually), to ensure that any changes in the repeatability or reproducibility of the measurement system are detected. The MSA should also be repeated in any of the following circumstances (where relevant):

- Before implementing SPC for the first time in a process – SPC charts will only be useful if the data being used to construct them is good quality.
- If a large process variation is seen and it is suspected to be caused by the measurement system
- After any significant changes or improvements to the process – if the performance of the process changes or improves significantly (for example if the tolerance limits change or the acceptance criteria change) a measurement system which was previously acceptable, may no longer be good enough – therefore the MSA must be repeated to check whether or not this is the case.
- After any significant changes in the people making the measurement – for example the involvement of new or less experienced people
- After a significant change in the measurement location or conditions – factors such as background noise (e.g. in an office), distractions, volume of work can all impact the repeatability and reproducibility of a measurement system.
- After any major maintenance of measurement equipment - especially if components have been adjusted or replaced.
- After measurement equipment has been moved – factors like temperature, humidity, excess vibration can all affect the performance of measurement equipment.
- As part of the acceptance for a new machine – don't assume that a new machine will have good repeatability and reproducibility
- As part of the training for new people to qualify that their inspection ability is comparable with the standard and with other members of the team

If you are unsure when or how often to repeat the MSA, discuss with a local Black Belt.

Repeating the MSA

After discussions with Alison, Julie and her team decide they will repeat their Gauge R&R every 3 months, until they gain confidence that the measurement system is performing consistently well. They update the relevant SOPs to state that the Gauge R&R must be repeated during the first week of every quarter.

After the first year, they find that the measurement system is performing consistently well, and after further discussions with Alison, they decide to reduce the frequency of the stability check to weekly, and that the Gauge R&R only needs to be done once per year, unless they have reason to suspect that there might be an issue with the measurement system.

Since the vernier calliper is calibrated every March, they decide that they will perform the Gauge R&R immediately after each annual calibration. This will enable them to check that the calibration hasn't affected the repeatability or reproducibility.

Continuous Data

Repeating the MSA

Since there are often changes in the people doing the in-line inspection (due to temporary staff being used when there are peaks in production volumes), Tom and his team decide that they will repeat their Attribute Agreement Analysis once per month. They are happy to do this, as it doesn't interfere significantly with the production process, and the analysis isn't too time consuming as the data can be imported from the production database into Minitab.

Although they find that the results of the Attribute Agreement Analysis are usually acceptable, the study occasionally highlights problems which they would otherwise have been unaware of, so they feel happy that they have chosen the correct frequency for the study.

Attribute Data

Summary

- A **Measurement System** is the combination of devices, hardware, software, procedures, standards, people and training involved in obtaining measurements

- If left unchecked there is a real risk that the measurement systems we use could be delivering unreliable data without us being aware of the existence of a problem. Therefore it is important to analyse how effective our critical measurement systems are.

- **Measurement System Analysis** is a structured procedure which we use to assess the ability of a measurement system to provide good quality data, before using it to collect data on which decisions are based.

- **Measurement System Analysis** provides a structured approach for teams to assess whether the measurement systems they use are **repeatable** and **reproducible**

- **Gauge R&R**, which is used when the measurement system is being used to collect **Continuous** data

- **Attribute Agreement Analysis**, which is used when the measurement system is being used to collect **Attribute** data.

Measurement System Analysis

How-to Guide

Change History					
Revision	Date	Description of Change	Author	Owner	Approval
V6.1	20/08/2013	Guide reformatted for SABRe	D Prodger	D Prodger	D Prodger

Document update policy

This document may be updated periodically. Major amendments will be shown as an update from one revision number to a higher revision number (e.g. revision 1 to revision 2) and therefore the content of the higher revision will be regarded as the latest requirements. A minor amendment will be shown as a number change after a decimal point (e.g. revision 1.1 to revision 1.2) and therefore any of these revisions may be regarded as the latest requirements until a major amendment is introduced