Calibration establishes traceability for the accuracy of a gauge. Gauge Repeatability and Reproducibility (GRR) establishes the ability for a gauge to accurately measure the desired tolerances in the environment in which the gauge is actually used.

As an example, you have a six inch metal shop scale that has been calibrated for accuracy. However, is this scale the right gauge for measuring the length of a part that has a tolerance of x.xxx? Likewise, would this be the correct gauge to use for measuring lengths of an interstate highway? Even though this scale has been calibrated, neither of these uses would be a good application for that instrument.

The Purpose of GRR studies

A competent GRR must take into account the tolerance that is being measured, the type of feature being measured, the operators taking the measurement, the conditions under which the measurement will be taken, the material being measured, the measurement technique itself, and other variables of inspection. Each of these points will be discussed in more detail below.

Ideally, the gauge being considered for any inspection process should have a GRR capability that is less than 10% of the specification tolerance being measured. If the GRR capability is between 10% and 20% of the tolerance being measured, management might be looking for an improved inspection technique but it will probably be a lower priority to other inspection issues (they probably have bigger fish to fry in the short term). If the GRR is more than 20% of the tolerance being measure, management needs to have a high priority for finding an improved measurement gauge, technique, or process. If the GRR is more than 10% of the specification tolerance, the user should get buy-in from the20% customer that the part measurement process is adequate to the customer's needs. The next section of will give an example of how GRR variation affects actual data collection.

When choosing any gauge, the feature being measured should influence the type of gauge chosen. Features of form such as flatness or roundness require gauges that can take continuous readings over the whole surface. A dial indicator that can traverse the surface will provide much more accurate information than a CMM machine that is only contacting 20 to 40 points. When choosing any gauge, there will always be tradeoffs between cost, gauge accuracy, and gauging process or technique plus other variables.

Any gauge process must also take into account the operators actually using the gauge. Do the operators have any special needs? I once had an excellent machine operator who was missing two fingers on one hand. He could not use common micrometers. The company came up with special holding fixtures for him so he could pass the GRR using the process that we modified for him at the time. Other operators at this operation did not object to using this fixture and they reported that it made their job easier.

What are the conditions in which the gauge is actually taking the measurements? Are the parts oily, gritty, or not yet deburred? Is shop air available to blow off the part prior to the measurement? Is the shop air filtered or will it have water mist in the air? Will the inspection take place in a quality lab or on the shop floor? I have noticed four distinct GRR conditions for most gauges. The best GRR values will be found when measuring parts at the gauge supplier's lab. The vendor technicians are experts at handling the gauge and they have developed repeatable "best-known" methods of using the gauges. I found that the most accurate GRR values came from having my metrology lab technicians use the gauge in my own metrology lab. This location is usually nearly ideal and my lab technicians were the most senior and experienced technicians that I had. There was a reduction in GRR when the gauge was taken to the shop floor but still operated by my metrology lab technicians. The shop floor was not temperature controlled, parts were more oily, there were more distractions to name just a few of the possible variables. The poorest GRR values came from the actual machine operators on the floor. These were the least experienced people using the gauge and the experience levels between operators could vary widely.

The material being measured can also affect the measurement process. Gauging for wood, plastic, or glass, may be different than gauging required for metals. Gauges for tolerance less than 0.002 inch might need to be air gauges no matter what the material. Will the shop floor process support the use of air gauges? Glass or ceramics materials might require non-contact gauges.

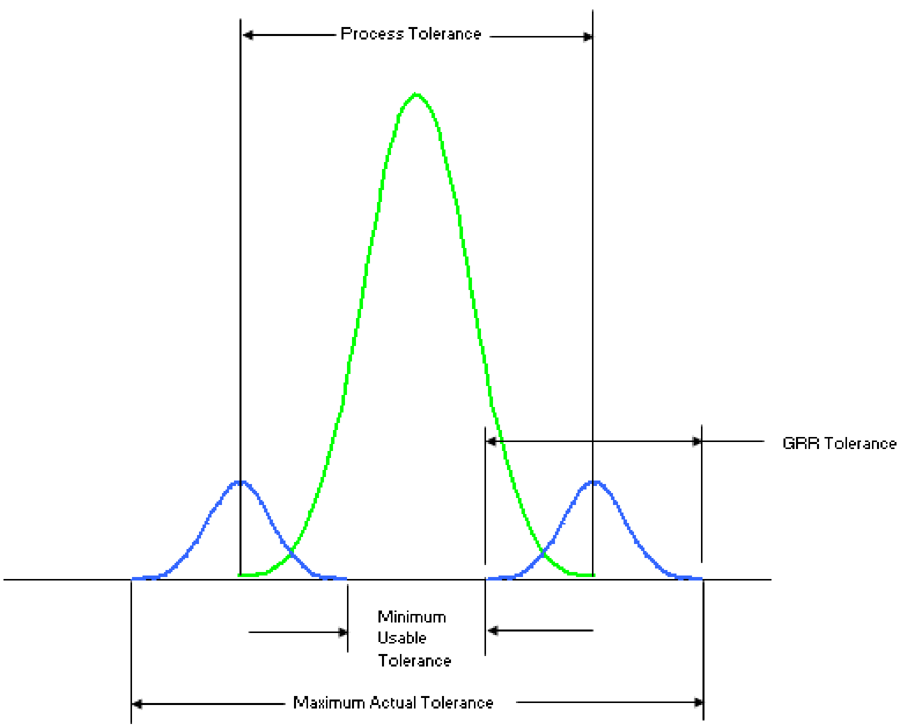
The measurement gauging process will have a major impact on the accuracy of the measurement technique. The gauging process must be firmly established and agreed to by all parties involved in the measurement. I recommend that a detailed flowchart be made of the process and that all operators involved agree to the process. Variations from the detailed flowchart could have major influences (either negative or positive) on measurements.

How often should GRRs be done? They should be done when establishing any inspection process. Any GRR should include a minimum of two operators. If a facility has one main inspection technician, who is the backup when that technician is out sick or on vacation? If any operator in the GRR is replaced, a new GRR should be done. A GRR should be done after each gauge calibration just to confirm that the GRR process is still valid.

Are their any occasions when one might not undertake a GRR? I have found that it is almost impossible for a process to be in control if the process has a GRR above 30% of the specification tolerance. If a process is plotted on an SPC chart and the initial data indicates that the process has a Cpk greater than 1.5, I would suggest that management might spend their immediate time on more pressing issues. At some point, a process might have a GRR run to fulfill overall quality objectives.

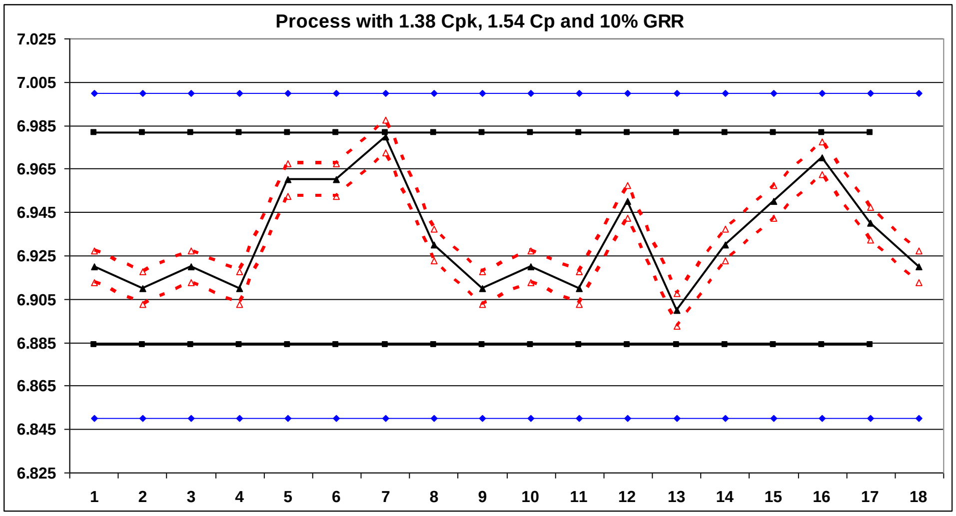
An Example of how GRR affects actual measured values

The affect of gauging variation on a measurement process is show in the following graph.
The actual process tolerance being inspected is the green line that represents a normal distribution. The variation in the inspection process (GRR) is represented by the two blue lines. I show both of these lines as normal distributions, each centered on the process tolerance limits. This graph shows that for any process and GRR combination, there is a Minimum Usable Tolerance that would need to be used if absolutely no bad parts were to forward from this operation. This graph also shows the Maximum Actual Tolerance possible due to the GRR variation that could result at the edges of the Process Tolerance.

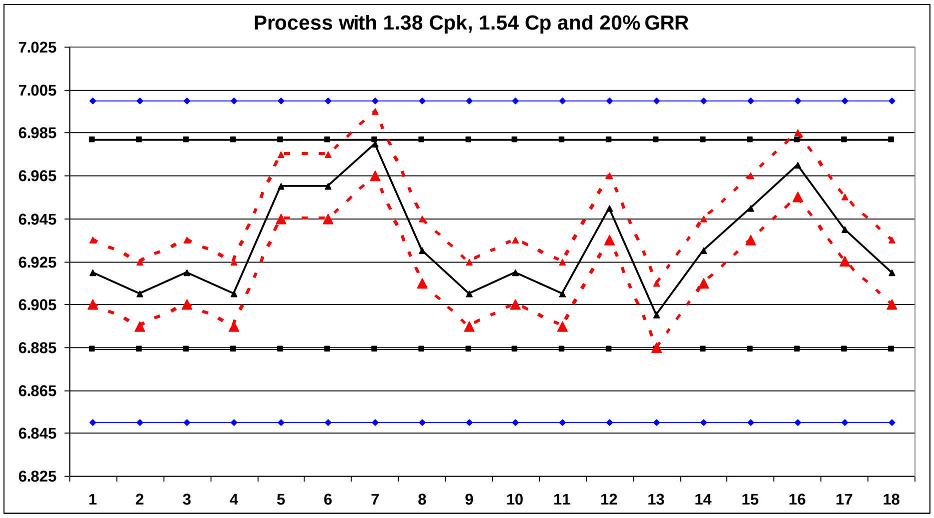


Below are three graphs showing how the variation in GRR affects the actual measurement of parts. In the three examples, The Cp of the data is a respectable 1.54. More importantly, the CpK of the sample data is 1.33 which means that the data is fairly well centered. In all three graphs, the tolerance limits are shown by the parallel blue lines. The process limits are shown by the two parallel black lines inside the blue tolerance limits. The actual measured data is shown by the solid black line. Parallel to the actual data line are two red dashed lines that represent the GRR variation as a percentage of the Process Tolerance.

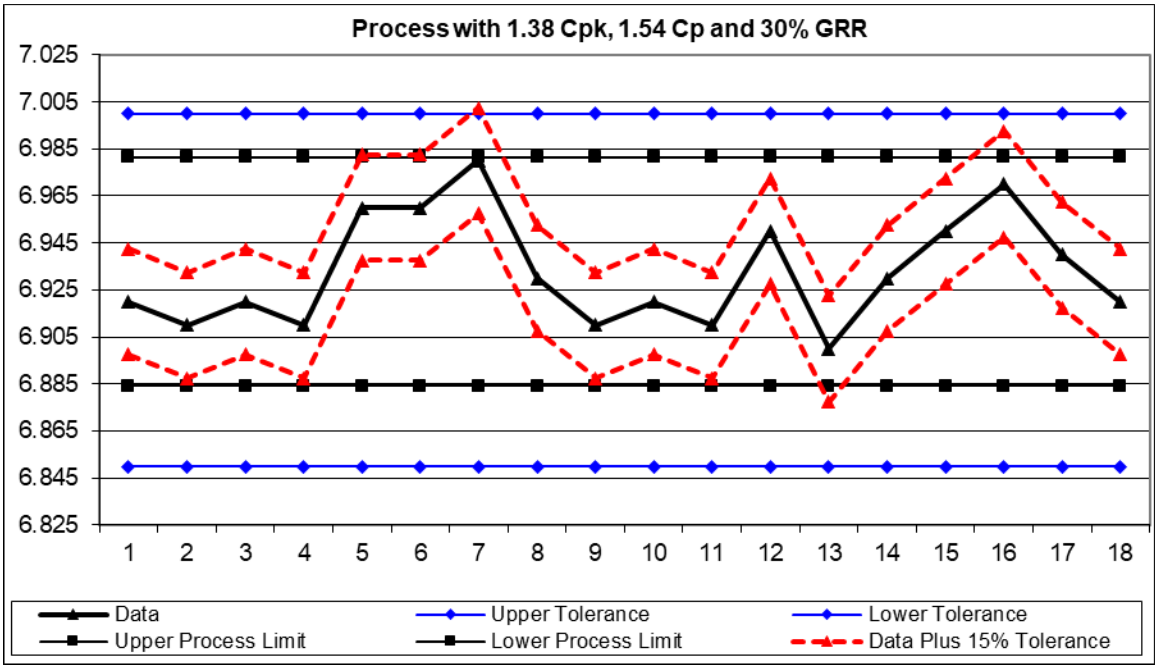
The first graph shows a GRR process that is 10% of the specification tolerance. The red dashed line shows how wide the actual data might vary based on the 10% tolerance being centered on the actual data taken. In this graph we see that one measured point is just inside the process limit. However, if the 10% process tolerance is centered on that one point, the actual measurement has a slight chance of actually being outside of the process limits.



The second graph shows a process where the GRR is 20% of the specification tolerance. This tolerance range is considered by most experts to be the largest allowable tolerance. With this expanded tolerance range, we see that two data points may actually exceed the process limits on the high side of the Process Limit and that an additional point may reach the Process Limit on the low side.



The third graph shows this same process but the GRR is now 30% of the specification tolerance. With this increased tolerance, four data points might actually exceed the upper process limit and an additional point might exceed the lower process limit. One point, in this case, is actually exceeding the upper specification limit and this is an unacceptable condition for anyone who is trying to insure that no "out-of-specification" parts are passed from this operation to an awaiting customer.



From these graphs, one can see that increasing the GRR as a percent of the specification tolerance may lead to acceptance or marginally bad parts or even rejecting marginally good parts. I can not recommend spending any more effort on a process if the GRR is less than 10% of the Tolerance Specification. I would think that a high priority be put on finding more accurate gauging or a better gauging process if the GRR is more than 30% of the tolerance specification. If the GRR is between 10% and 30% of the tolerance specification, I suggest that the inspection department pursue looking for a more accurate gauge and better gauging processes. The urgency of this pursuit will depend on how critical this feature is and other urgencies that the department might be facing at the time. Obviously, a GRR of 12% is much less urgent than a GRR of 28%. If the GRR is over 40% of the specification tolerance, I recommend that the quality department could save money on gauging and just flip a coin whenever they need to inspect that feature.

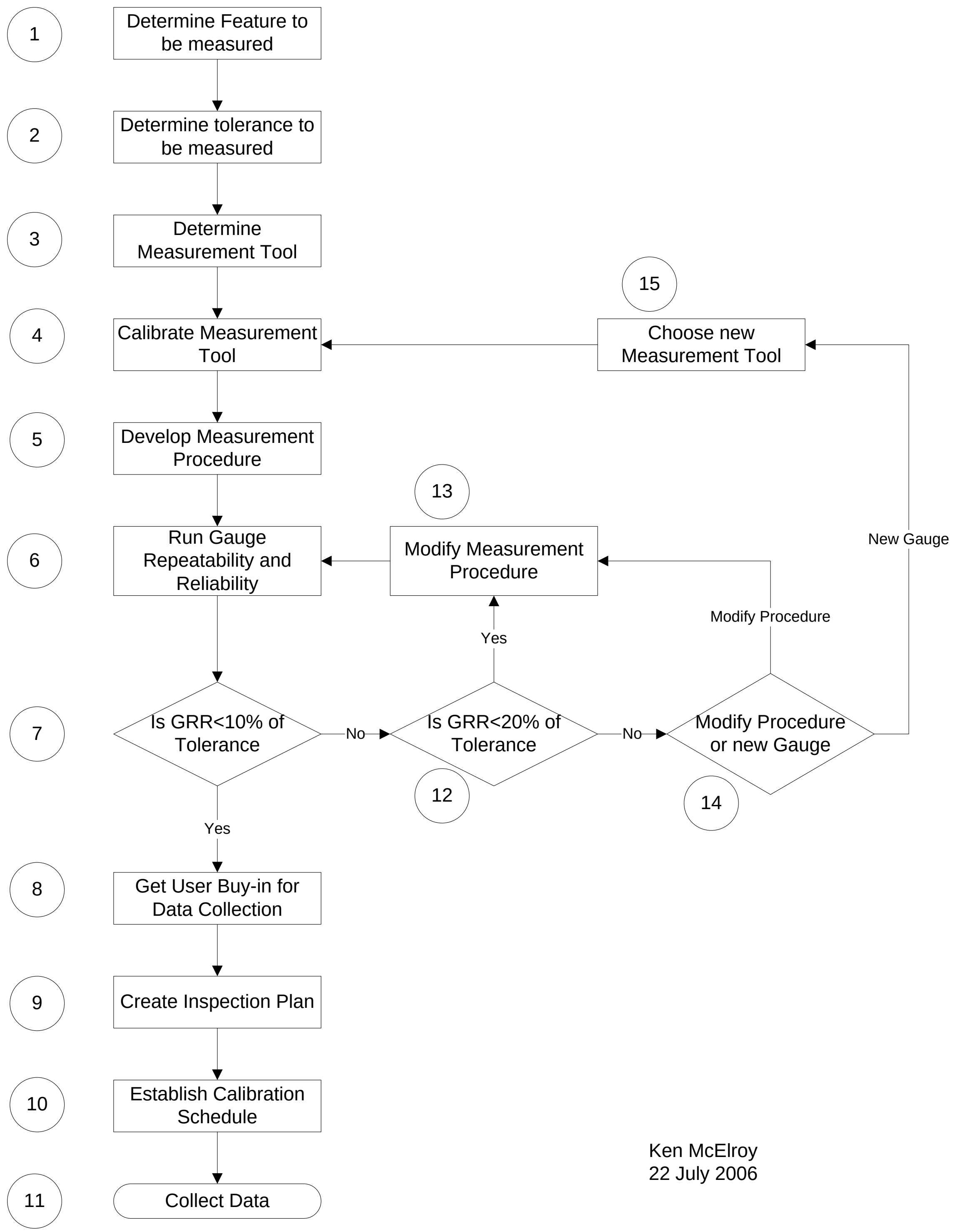
When Should a Gauge Repeatability and Reproducibility Procedure be Used?

If you are opening a new manufacturing line or facility, you should consider doing a GRR for every dimension that you will be checking. Your manufacturing facility will thus start off on the highest level. However, if you are introducing SPC and GRR into an existing facility they you will have to prioritize you assets. If you perform a random SPC check and your Cpk is larger than 1.5, I would suggest you move to another feature and document your GRR on this feature as time and assets allow. If your SPC check found your feature out of statistical control, a GRR would be required before your start improving your Cp on that feature so you could determine how much of your variation is from the gauge and how much is from the process.

Once a gauge has passed a GRR for the application in which it will be used, does the GRR ever need to be redone? My experience suggests that a GRR should be redone if you change an operator at the application in question. Each new operator may have a vastly different background from existing operators and a GRR would help determine if the new operator needs extra training, has special needs, or brings a new technique to using the tool that may be shared with the older operators at that application. I have also found evidence that a GRR should be done after the gauge is calibrated. Repeating the GRR after each calibration confirms that the GRR process is still being used correctly.

Basic Steps in Performing a Gauge Reliability and Repeatability Procedure

The first step in any Gauge Reliability and Repeatability study is to determine what feature is to be measured. The feature may be a critical feature on a print that requires SPC capability in production. The feature may be one that has been determined to be process critical in order to make the finished part. The type of feature chosen will determine what type and size of inspection equipment is required.



The feature chosen for inspection should have a tolerance called out on the print or specification. This tolerance will to a large degree determine the cost of the inspection equipment. If the tolerance is +/- 0.020 or a surface finish of 125 or a circularity of 0.020, the required accuracy and cost of the inspection equipment will be much less than if the tolerances are +/- 0.002 or a 8 surface finish or a

0.002 circularity. Gauging that will be used on the shop floor must be more robust than equipment used in an inspection lab. Gauging requirements will also vary depending on volume, frequency of use, required versatility, of if it is a short run production.

Once the feature and the tolerance are known, an inspection tool may be chosen. Remember that an instrument that looks good on the show room floor does not always perform adequately on the shop floor in your specific environment. Choosing an instrument that measures a length of +/- 0.005 is easy but choosing an instrument that measures a +/- 0.5 angle over a 0.003-4 length will be exceedingly difficult to obtain.

Once the inspection gauge has been obtained, it must be calibrated. This step is beyond the scope of this paper and the procedures are well established.

After calibration, a measurement procedure must be developed. This procedure must be documented to include written instructions. I have found a flow chart with embedded notes to work in most cases. This procedure should include elements on gauge setup, how and when to pick parts for inspection, cleaning and holding or positioning the parts during the inspection, and where and how to record the inspection data. There may be a need for experimentation before this document is completed. During this step, the same parts may be measured several times to see if the measured variation is within allowable tolerance limits.

Once the inspection department feels that they have an adequate measurement process, the initial Gauge Repeatability and Reproducibility study may be made. This study should involve a minimum of two operators measuring several parts a minimum of two times. The operators should be the operators that will be measuring the parts in actual production. If the facility is operating on three shifts, then all three operators should take part in the GRR study. All operators will measure the same parts at least twice and preferably three times. They should measure the parts in random order. It is important that this study be done on the shop floor or where ever the production parts will actually be measured. A GRR done in the inspection lab will always show better results than a GRR done on a dirty shop floor.

Once the GRR study has been finalized, it needs to be institutionalized. The people actually making the production inspections must agree to the inspection process used in the GRR study. If they do not agree, then the inspection process must be modified and the GRR study redone until these people are happy with a process that results in an acceptable GRR. If the people making the study do not agree with the inspection techniques, they will start to cut corners or adopt their own technique which may introduce widespread and unacceptable variations into the inspection process.

It is now time to create the formal inspection plan. This is similar to the measurement procedure but is a formal document. A flowchart and other written notes are required. This is a quality document that should be filed so it can be reviewed from time to time with personnel inspecting the parts in production. These notes may also be used when training new inspectors or operators.

Don't forget to establish a calibration schedule for the inspection equipment. This step is usually required by ISO or other quality systems. I have found it a good idea to review the GRR of any piece of equipment when it is time for its periodic calibration, usually yearly. This is a good time to rerun a GRR to make sure that the process is still being followed and that the current operators are still properly trained.

Now, let is step back to the results of the GRR. What if the results of GRR were greater than 10%. If the result was between 10 and 20% of the Specification Tolerance, I would first try to modify the inspection process to see if the GRR could not be improved. One frequent cause of GRR failures is variation within production parts. This normal variation may be reduced per instructions on how to position a part during the GRR study itself. However, the whole purpose of the GRR study is to implement SPC studies that will reduce this variation so this variation should be expected in the beginning. If part variation is thought to be a problem, the GRR study might be run using gauge pins, gauge blocks, or other precision parts. If the GRR study is found to be adequate using these precision parts, then the GRR study is probably adequate and it should be formalized. The SPC process will then begin to resolve variations within the production methods and future GRR studies might use the parts from the improved process.

If the GRR using precision parts is not acceptable, and modifying the inspection process does not produce an acceptable GRR, than the inspection department should look for a more precise or robust gauge. This may not be an easy task to complete but it should be a high priority.

Most Common Excuses for Failing a GRR
Basic GRR Example

One of the most common excuses for failing a GRR is that the inspection technique is not consistent. This is why the step about documenting the inspection process prior to making the GRR study is so important. It is also because of this complaint that operator buy-in is important.

Another common excuse is that the operators were not properly trained. Once again, proper discussion about the inspection process with the operators prior to the GRR will void this excuse. I find this excuse common in shops where the inspection department assumes that all of the operators are adequately trained and that any "man off the street" can inspect a part.

The parts are not consistent is another excuse. This is probably a good excuse since the GRR is the first step in developing an SPC process whose purpose is to reduce variation. However, to bypass this excuse, use precision gauges or parts to do a GRR study. If the GRR passes using the precision parts, it confirms the suspicion that actual part variation caused the original GRR failure. If the GRR study is acceptable using production parts, the results of the initial SPC study will probably be pleasing.

The ever popular excuse that the gauge is not adequate also has some merit in many cases. If the gauge can not pass the GRR using precision parts and after managing the inspection process, then it should be saved for less rigorous inspection duties and a new, more precise gauge needs to be pursued.

What are the different types of GRRs

GRR comparisons usually fall into three categories. The original method was to compare the results of using various gauges by statistical differences in the various means and standard deviations. This is the basic Null Hypothesis for the average and standard deviation of two or more gauges or two or more inspectors. This approach also relies on graphical comparisons which is popular with many people.

Prior to computers, several approaches involving statistical methods were created. The General Motors Long Form technique was taken up by the Automotive Industry Action Group (AIAG) with minor modifications. This requires simple math and could be used with up to ten samples, four operators, and three trials for each operator. More details on this technique can be found in the AIAG publication, Measurement Systems Analysis. Larry B. Barrentine presents a popular variation on this approach in his Concepts for R&R Studies. Mr Barrentine points out that if the number of operators times the number of samples is less than 15, then special constants need to be used. However, he does not mention the Part Variation calculation presented in the AIAG publication. The main advantage of these methods is simple graphs can be created that show how the variables change between operators and parts.

With the advent of computers, the Analysis of Variance (ANOVA) technique has become popular. This approach gives the answers in terms statisticians love. However, I have found that most people are intimidated by the ANOVA and these people, especially managers not trained in statistics, will bypass this approach if they can see the graphs from the older method.

For this class, I will use a variation of the AIAG method with graphs comparing the operators and the parts themselves.

A Sample Problem as a GRR Class

For this class, I have chosen to measure pennies. These are readily available. Each penny for this class is from a different year so we can identify each penny for recording measurements. I chose six pennies, and not more, so the calculations at the end will not be too involved.

I have also chosen three instruments to compare. I have a wooden Vernier caliper I chose over a $1.00 plastic Vernier caliper since most people can not read these old Vernier scales anymore. I have also chosen a 1 inch Starrett mechanical micrometer serial number 436 that has a ratchet stop thimble for higher sensitivity. My third instrument is a electronic Vernier micrometer, Mitutoyo Digimatic Model CD-8"P, Serial Number 7008914 that reads to four places.

For this class, the participants may be broken into three groups, one for each instrument. I let the managers or least experienced people use the wooden Vernier caliper. The pennies can then be passed to each group which will measure each penny twice. I could use three measurements but I chose two measurements to save time. The readings should be made as random as possible. It usually works well for one person in each group to measure the pennies while another person records the values and then they switch functions for the second set of readings.

Once all of the groups have measure all six pennies, the class is pulled back together and the data entered into a GRR form. A short form just for this class is shown below. For this class, I will assume that the tolerance on the outer diameter of the pennies is +/- 0.010 inch. I haven't really checked with the US Mint to see what their actual tolerance is, but this is only an example.

Discussion of Results of the Basic GRR Example

Below is a chart that gives the results of two operators measuring each of the six pennys twice.

	Wooden Vernier Calipher		Micrometer		Electronic Vernier Calipher	
	Try 1	Try 2	Try 1	Try 2	Try 1	Try 2
	Operator A		Operator A		Operator A	
1969	24/32 = .75	24/32 = .75	0.7500	0.7500	0.7475	0.7480
1978	24/32 = .75	24/32 = .75	0.7520	0.7520	0.7505	0.7520
1979	24/32 = .75	24/32 = .75	0.7500	0.7490	0.7500	0.7495
1980	24/32 = .75	24/32 = .75	0.7480	0.7480	0.7475	0.7470
1986	24/32 = .75	24/32 = .75	0.7490	0.7490	0.7490	0.7495
1993	24/32 = .75	24/32 = .75	0.7500	0.7510	0.7495	0.7505
	Operator B		Operator B		Operator B	
1969	24/32 = .75	24/32 = .75	0.7500	0.7500	0.7485	0.7475
1978	24/32 = .75	24/32 = .75	0.7520	0.7520	0.7505	0.7510

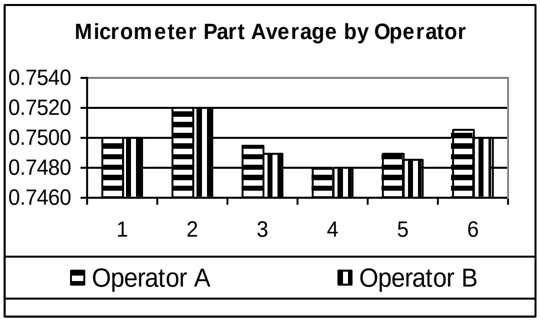
1979	**24/32 = .75**	**24/32 = .75**	**0.7490**	**0.7490**	**0.7490**	**0.7495**
1980	**24/32 = .75**	**24/32 = .75**	**0.7480**	**0.7480**	**0.7480**	**0.7470**
1986	**24/32 = .75**	**24/32 = .75**	**0.7480**	**0.7490**	**0.7485**	**0.7485**
1993	**24/32 = .75**	**24/32 = .75**	**0.7500**	**0.7500**	**0.7490**	**0.7500**

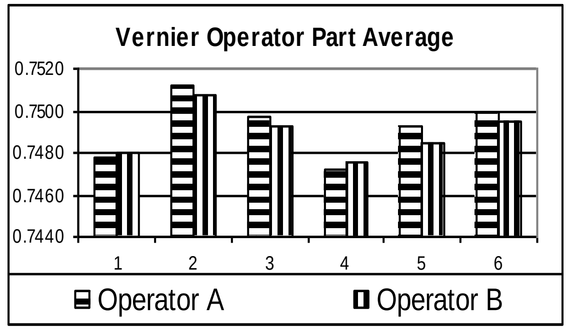
The first thing to notice is the difference in accuracy between the three measurement instruments. The wooden Vernier calipher is divided into units of 1/32 on an inch. The pennies all measured 24/32 which I convert to the more familiar decimal 0.75 inch. The Micrometer is measuring to 0.001 inch and many people will interpolate to an extra 0.0005. The electronic Vernier calipher reads to an accuracy of 0.0001 and because of the four digits, it shows the most variation of these three instruments.

I show the summary of the calculations below. Don't worry about the actual calculations. Various software packages can do this for you or you can go to the magazine website where I have loaded a free Excel file that will make these calculations for you. Note that all of the calculations for the wooden calipher are zero. From these calculations, we can see that this instrument does not have the accuracy to measure the pennies. Using this instrument would be akin to using the six inch scale to measure the diameter of these pennies. I will not discuss the wooden calipher any further.

	Wooden Calipher	Micrometer	Electronic Vernier Caliper
REPEATABILITY(E.V.)	0.000	0.0011	0.0032
REPRODUCIBILITY(A.V.)	0.000	0.0009	0.0005
GAUGE R&R	0.000	0.0014	0.0033
PART VARIATION	0.000	0.0077	0.0070
TOTAL VARIATION	0.000	0.0078	0.0077
Recommended Process Tolerance	0.000	0.0142	0.0327
Minimum Recommended Process Tolerance	0.000	0.0071	0.0164

Repeatability in these calculations is the measurement of equipment variation. From the above comparison, note that the micrometer actually has less repeatability than the electronic Vernier caliper. I have noticed that in nearly all of the GRR classes that I have given, the micrometer is better in repeatability. I have noticed that this gauge gives much better GRR values if the ratchet stop thimble is used when clamping the pennies.





Reproducibility is a function of operator variation. In our sample, the electronic Vernier calipher shows less reproducibility than does the micrometer which means that the two operators were more consistent using the electronic Vernier than using the micrometer. The micrometer showed variation on parts 3, 5, and 6. The Verier showed variation on every part. The range of variation with either gauge was 0.004.

Gauge R&R is the vector sum of the Repeatability and the Reproducibility. Because of the much lower value in Repeatability, the micrometer shows up better overall. Gauge R&R determines the tolerance that this instrument should be used for. Ideally, the Gauge R&R should be less than 10% of the tolerance being measured so a tolerance 10 times the Gauge R&R would apply. This calculates out to a tolerance of 0.014 for the micrometer but 0.0327 for the Electronic Vernier Caliper. This means that the micrometer would be acceptable for the 0.020 tolerance in this example. Many companies feel that a the Gauge R&R may be up to 20% of the tolerance being measured. In this case, the tolerance of 5 times the Gauge R&R would apply. This calculates out to 0.007 for the micrometer and 0.01635 for the electronic Vernier. These calculated values are half of the value calculated for 10% Gauge R&R and I don't recommend their use if more accurate gauges can be found. If we accept 20% GRR, either gauge would be acceptable for the 0.020 tolerance.

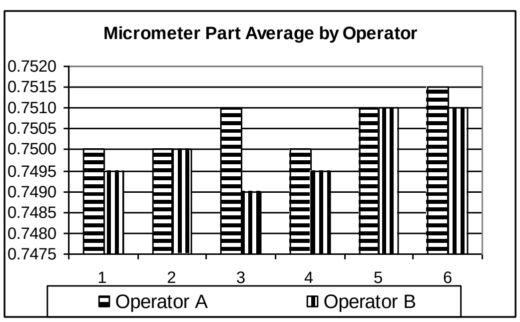
The summary also shows the part variation for these pennies. Both the micrometer and the Electronic Caliper found that the part variation was over 0.007. This is over twice the GRR value for either gauge and is very significant. Should a gauge fail a Gauge R&R when the Part Variation is large, I would recommend that the Gauge R&R be rerun using gauge pins or gauge blocks. Using these pennies, I could go look at the pennies 3, 5, and 6 since both instruments showed variation on all three of these pennies.

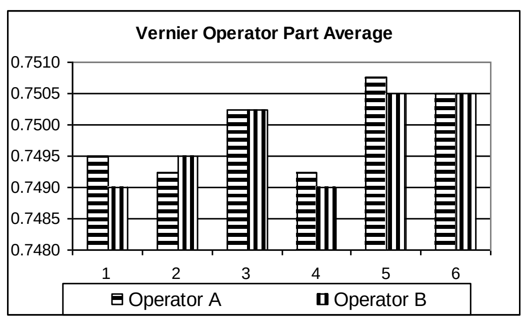
Discussion on GRR using Gauge Pins

Below is a chart for data taken using six new pennies. Each penny is identified by a different year to keep them separated. In this study, the measurement procedure had also been modified. Each measurement was taken with the portrait of Lincoln in an upright position. In the previous study, no mention was made to the orientation of the pennies during measurements.

	Micrometer		Electronic Vernier Calipher	
	Try 1	Try 2	Try 1	Try 2
	Operator A		Operator A	
2000	0.750	0.750	0.7495	0.7495
2001	0.750	0.750	0.7490	0.7495
2002	0.751	0.751	0.7505	0.7500
2003	0.750	0.750	0.7490	0.7495
2004	0.751	0.751	0.7510	0.7505
2005	0.752	0.751	0.7505	0.7505
	Operator B		Operator B	
2000	0.749	0.750	0.7490	0.7490
2001	0.750	0.750	0.7495	0.7495
2002	0.749	0.749	0.7500	0.7505
2003	0.750	0.749	0.7490	0.7490
2004	0.751	0.751	0.7505	0.7505
2005	0.751	0.751	0.7505	0.7505

The raw data for each penny is given in the above chart. Once again, two operators measured each of the pennies twice in a random order.





The first thing that I noticed from this data is the Micrometer had one reading that varied by 0.002. The total range variation for the Micrometer is 0.003. The readings from the Vernier look much closer on this sample. No penny varied more than 0.0005 between the four readings of the two operators. The total range variation for the Micrometer over all six pennies was 0.002.

	Micrometer	Vernier
REPEATABILITY(E.V.)	0.0011	0.0010
REPRODUCIBILITY(A.V.)	0.0021	0.0004
GAGE R&R	0.0024	0.0010

PART VARIATION	0.0024	0.0029
TOTAL VARIATION	0.0034	0.0031
Recommended Process Tolerance	0.0239	0.0102
Minimum Recommended Process Tolerance	0.0120	0.0051

Comparing the results of these GRRs, the Micrometer actually calculates nearly twice as badly using the gauge pins as was calculated in the first GRR. This is interesting since in both cases, three pennies were found to vary by 0.001. This variation represents the limit that this instrument may practically be used and shows that because of measurement variations, no two GRR tests will agree perfectly.

The Vernier however showed marked improvement when measuring the gauge pins over measuring the pennies. The recommended process tolerance with the original pennies was 0.3277 inch while it was only 0.0102 with the gauge pin pennies. We now have confidence that the Vernier is indeed a more accurate gauge and we can use it to check the original pennies in detail.

The class remeasured the original pennies in detail using the electronic Vernier and they noted that all of the pennies had flat spots on the outer diameter where I had filed them to exaggerate their variation.

Summary

Even though the Micrometer is a much older instrument than the electronic Vernier caliper, the original GRR shows that it is and adequate instrument to use for measuring the outer diameter of pennies that have the tolerance used in this study. However, we note that the part variation on these original pennies is so large that it is affecting the Gauge R&R and the Gauge R&R value, especially on the electronic caliper. Once the GRR study was repeated on gauge pins, the Vernier proved to be more accurate. If our goal is merely to monitor the penny's outer diameter, then either gauge would work. If our goal is to reduce the variation or tolerance in the penny's outer diameter, then the electronic Vernier would be the better gauge of these two.

Gauge Repeatability and Reproducibility studies measure the accuracy of a gauge in a given application, under actual working conditions. The Gauge R&R calculation may be affected by the inspection process, by the inspectors themselves, and even by variations within the parts being used in the study. Should a gauge fail a Gauge R&R study, each of these variables must be reviewed in detail.