Looking beyond audit-oriented evaluation of gauge repeatability and reproducibility: A case study

TIRTHANKAR DASGUPTA & S. V. S. N. MURTHY
SQC & OR Unit, Indian Statistical Institute, 98 Sampatrao Colony, Baroda 390007, Gujarat, India

ABSTRACT *With the advent of QS 9000 quality system, measurement systems analysis (MSA) has begun to get its due importance. Companies seeking QS 9000 certification have been conducting gauge repeatability and reproducibility (R&R) studies regularly. Such exercises are proving to be immensely beneficial from the point of view of in-depth understanding of the company's own measurement process. However, the common Indian experience is that most of the industries conducting MSA studies are finding it extremely difficult to investigate the root causes behind poor gauge R&R and consequently to take proper corrective actions. Most of the exercises are thus ending solely in an audit-oriented evaluation of the gauge R&R without any systematic e ort to reduce the same. In this paper, we have made an attempt to demonstrate through a case study how the problem of high gauge R&R can be resolved by unearthing the root causes and taking appropriate corrective actions.*

Introduction

Gauge repeatability and reproducibility (R&R) studies have become a mandatory requirement for all companies seeking the QS 9000 certification all across the globe. These studies are now religiously being conducted in such industries; involving plenty of data generation, training on statistical methods and analysis of data using statistical software. Unfortunately on many occasions the study is being restricted to a simple evaluation of the gauge R&R with the objective of satisfying third-party auditors and no actions following the same. There are also instances where unsatisfactory values of gauge R&R are leading to indiscriminant change of the measuring device without any actual investigation. Such actions involving additional investment need not necessarily improve the gauge R&R since the actual trouble may lie elsewhere. Thus, meaningless evaluation of gauge R&R is creating an increased level of frustration among people who are conducting these studies and considering the complications involved in conducting such studies, one big and pertinent question often being raised is "Are these studies really useful?" The case study described here is expected to provide a strong affirmative answer to this question.

This study was conducted in the radiator division of a company manufacturing radiators and gaskets for automobiles. Four cases where the percentage gauge R&R was very high are

Correspondence: T. Dasgupta. E-mail: isi brd@satyam.net.in

briefly discussed. In all of these four cases the underlying causes leading to high measurement variation were identified and appropriate corrective actions helped in significant reduction of the gauge R&R. It will be interesting to note that in each case the nature of the root cause and the subsequent corrective action were different from the others. Out of the four cases, only one situation demanded a change in the measuring device.

We begin with a brief description of the methodology for conducting gauge R&R study.

Repeatability, reproducibility and conducting gauge R&R studies

According to the QS 9000 reference manual on measurement systems analysis (MSA), repeatability is the variation in measurements obtained while measuring a given characteristic repeatedly on the same part with one measuring instrument by the same appraiser. Reproducibility is the variation in the average of the measurements made by different appraisers using the same measuring instrument on identical characteristic of the same part.

Usually, 'gauge R&R' studies are conducted so that repeatability and reproducibility can be estimated from a single experiment.

A number of appraisers, say n, are selected at random from the set of people who normally take the measurement under consideration. A number of parts, say m, selected randomly from the process are measured at least twice by each appraiser. Although the analysis of variance (ANOVA) technique can be used for analysing the data, mostly the method of control charts (for average and range) is used to determine whether the measurement system is under control and evaluate the gauge standard deviation s_0 and the true appraiser standard deviation s_1. Repeatability or equipment variation (EV) and reproducibility or appraiser variation (AV) are, respectively, given by

$$EV = 5.15 * s_0 \text{ and } AV = 5.15 * s_1$$

The 'gauge R&R' is given by $\sqrt{(EV^2 + AV^2)}$.

The total variation (TV) which comprises EV, AV and part-to-part variation (PV) is calculated from the set of measurements and is given by

$$TV = \sqrt{(EV^2 + AV^2 + PV^2)}$$

The gauge R&R is usually expressed as a percentage of TV. As per the QS 9000 guidelines, a measurement system having more than 30% gauge R&R is considered inadequate and calls for immediate corrective actions.

Preliminary gauge R&R studies

From the point of view of criticality, four critical characteristics of radiator components were chosen for 'gauge R&R' study, viz. *Filler neck hole dia*, *Tube width*, *Tube crown thickness* and *Inlet hole dia*.

For each of these characteristics, sample parts were chosen from manufactured lots. Each part was measured thrice by three appraisers. The parts and sequence of measurement were randomized as a safeguard against bias in measurements. The data were analysed using the method mentioned in the previous section. Some important aspects of the four measurement systems (gauge used, least count, material of part measured) and results of the preliminary analysis are summarized in Table 1.

As seen from the table, all the measurement systems were thus found to be inadequate. Hence, each of these measurement systems were studied in a systematic way to unearth the root causes resulting in such high R&R values.

Table 1. *Results of preliminary study*

Characteristic	Gauge	Least count (mm)	Material of part	*TV*	*%EV*	*%AV*	*%R&R*
Filler neck hole dia	Vernier calipers	0.01	Brass 0.8 mm	0.1107	35.77	72.88	81.12
Tube crown thickness	Dial vernier calipers	0.02	Brass 0.12 mm	0.1866	49.19	38.06	62.10
Tube width	Dial vernier calipers	0.02	Brass 0.12 mm	0.0365	10.71	43.83	46.87
Inlet hole dia	Vernier calipers	0.01	Brass 0.8 mm	0.1057	86.56	20.09	88.83

Investigation of root causes

Filler neck hole dia

For this characteristic, both the repeatability (35.77%) and the reproducibility (72.88%) were found to be high. The extremely high value of reproducibility was suggestive of some special cause resulting in appraiser-to-appraiser difference in measurements. The appraisers' average chart (Fig. 1) was plotted to identify the suspected systematic pattern in the measurements taken by the three appraisers. The chart revealed that appraiser C had a tendency to measure the parts on the lower side, whereas appraiser A had a tendency to measure the parts on the higher side.

This systematic pattern immediately led to the root cause, which was identified by the appraisers themselves, having been given the feedback. The filler neck holes were slightly tapered. Since no markings were provided on the gauge during the study, the difference in amount of insertion of gauge into the filler neck hole between the appraisers resulted in high *%AV* (see Fig. 2). This cause was also likely to affect the repeatability to a certain extent due to variation in the amount of insertion of gauge by the same appraiser.

The gauge used for measuring this dimension was marked on its jaws, which would give an indication of how much the gauge was to be inserted while taking the measurement.

A fresh study conducted with the marked gauge yielded the following results:

$$\%EV = 17.53,\ \%AV = 14.03,\ \%R\&R = 22.45$$

Thus, the gauge R&R was substantially improved after implementation of corrective actions.

As a bi-product of this study, the taper in the filler neck hole was detected and efforts were undertaken to reduce the same. Thus, this exercise not only helped in improvement of the measurement system, but also unearthed a hidden quality problem.

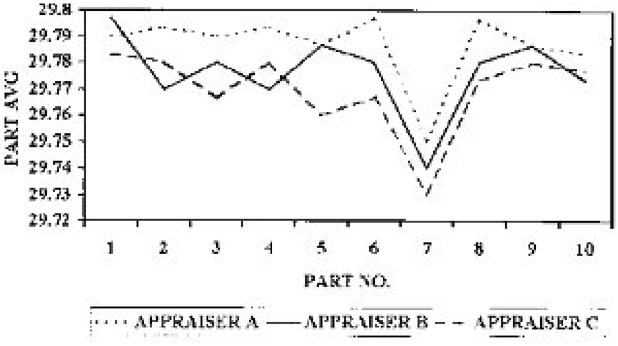

Figure 1. *Appraiser average chart.*

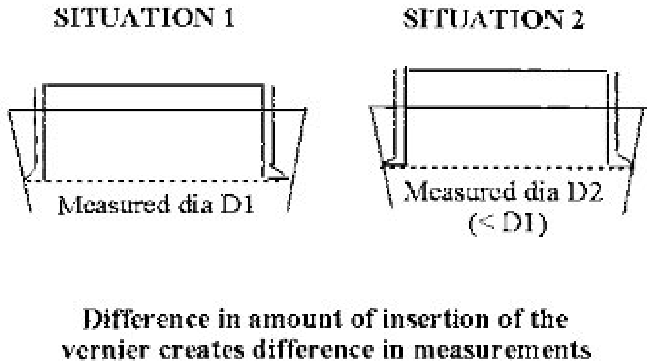

Figure 2. *E ect of taper on measurement of Filler neck hole dia.*

Tube crown thickness

For this dimension, both repeatability and reproducibility were found to be high but not significantly different. No systematic difference among the appraisers could be detected by the appraisers' average chart. The gauge was also found to be in proper condition. However, after a close look at the measurement method during the study and after discussing with the appraisers, it was highly suspected that the variation in force applied by the same appraiser at different times and the variation in force applied among the appraisers on the flexible tube were causing both the repeatability and the reproducibility to be greater. The tube made of brass was only 0.12 mm thick and a slight application of force was likely to result in a smaller measured value. The suitability of the measuring equipment, i.e. dial vernier calipers in measuring the above characteristic, was thus questioned.

In order to validate the above suspicion, a fresh study was conducted by replacing the dial vernier calipers with a dial thickness gauge which required no application of force at the time of measurement. The results of analysis were as follows:

$$\%EV = 20.48,\ \%AV = 0,\ \%R\&R = 20.48$$

The change in measuring equipment made the measurement system acceptable and thus validated the above suspicion.

Tube width

In this case, the reproducibility was found to be high. However, no striking differences among the appraisers were observed. Although the component (tube) and hence the material thickness were the same as in the previous case, unlike the crown thickness the effect of variation in force applied could not have resulted in a large reproducibility owing to the basic nature of the dimension. A glance at Fig. 3 will help in understanding this aspect. The tube at the width side was much stiffer compared with that at the crown side due to the presence of folds and smaller contact area. The location at which the measurement was to be taken was marked on each tube, thereby ruling out the possibility of appraiser-to-appraiser variation on that account. The calibrations on the gauge dial were sufficiently clear.

Thus, all the possibilities which could have resulted in high reproducibility (as suggested in the QS 9000 MSA reference manual) were explored one by one and yet no concrete direction could be found. This gave rise to the suspicion that it might not be a large value of

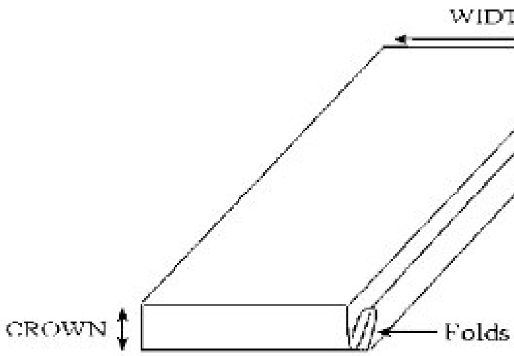

Figure 3. *Crown and width of a tube.*

AV, but a small value of *TV* which was giving a misleading picture of the measurement system variation in terms of percentage *AV*.

A careful look at the computations showed that the total variation (*TV*) as calculated from the generated data was 0.0365. However, through a statistical process control (SPC) study conducted earlier, the capability of the process (six times the inherent process standard deviation) was calculated as 0.1631. This anomaly suggested that there were possibly some drawbacks in the gauge R&R exercise itself. The QS 9000 MSA manual (chapter II, section 3, p. 39) clearly states that "The sample parts must be selected from the process and represent its entire operating range. This is sometimes done by taking one sample part per day for several days. This is necessary because the parts will be treated in the analysis as if they represent the full range of product variation that exists in production". It was thus evident that the sample parts selected for study were by no means representative of the process.

The MSA reference manual states that "If the process variation is known and its value is based on 6σ, then it can be used in place of the total study variation (*TV*) calculated from the gauge study data". Since the value of 6σ was obtained earlier as 0.1631, for computation of % gauge R&R, the denominator was taken as $TV = 5.15\sigma = 5.15 \times (0.1631/6) = 0.14$. Thus, the revised values of percentage *EV*, *AV* and gauge R&R were, respectively, given by

$$EV = 2.79\%, AV = 11.43\%, R\&R = 12.22\%$$

Thus, in this particular case, the gauge R&R study had not been carried out in the proper way. This resulted in an overestimation of the measurement variation and gave a false impression that the measurement system was highly inadequate.

Inlet hole dia

For this dimension, it was observed that the repeatability was extremely high (86.56%). The history of the gauge suggested that it was in good condition; hence maintenance of the gauge was not a problem. However, studying the characteristic inlet hole dia carefully, it was discovered that the major reason was excessive within-part variation due to ovality. The QS 9000 MSA manual clearly states that "within-part variation, such as taper or out-of-round can cause the measurement system evaluation to provide misleading results. This is because unaccounted within part variation affects the estimate of repeatability, reproducibility, or both". In this case, the high ovality created problems for the appraisers to locate the diameter properly.

The QS 9000 manual suggests a reformulation of the MSA study under such circumstances. A repeat study was thus conducted in which the diameter was measured in various directions to span the full range of variation for each hole. The maximum, minimum and range of the measured diameters for each hole were recorded. Each part went through this process three times by each appraiser. This helped in partitioning of the total variation into four components - *EV*, *AV*, *PV* and within-part variation. It was observed that the percentage *EV*, *AV* and within-part variation (*WIV*) were, respectively, given by

$$\%EV = 20.04,\ \%AV = 8.98,\ R\&R = 21.97 \text{ and } \%WIV = 86.09$$

Thus, the measurement system was adequate, but excessive within-part variation had resulted in a high gauge R&R earlier. However, it was evident that the existence of high ovality was a matter of concern to the company and major focus should be given on the manufacturing process to reduce this undesirably high value of ovality. A team-oriented problem-solving approach was employed to identify the root cause of ovality and it was discovered that ovality was due to directional stress problems. Accordingly, the cutting tool was modified by providing allowance on its dimensions to reduce ovality. Thus, similar to the case of *Filler neck hole dia*, this particular study also helped the company to unearth a hidden quality problem and, accordingly, initiate corrective actions.

Concluding remarks

The success achieved in the above four cases resulted in twofold benefits: apart from reducing the measurement variation, it boosted the confidence of the people regarding their ability to conduct gauge R&R studies purposefully and effectively. Although further reduction of % gauge R&R is desirable in all the four cases (gauge R&R below 10% is considered to be ideal) the current achievement was considered to be a pioneering effort in this particular company and paved the way for further successful studies. It could be established that the belief 'gauge R&R studies are meant for academic interest only' was a myth.

It is clear that in the case of tube width, proper planning for the MSA study (more specifically, right choice of the sample parts) would have given us the correct results straightaway. It is important that people conducting MSA studies follow each step suggested in chapter II, section 3 (Preparation for a measurement system study) of the MSA reference manual meticulously to avoid confusion at a later stage.

Considering the problems encountered during this study, it was felt that it would be convenient if some generic guidelines for investigating the causes behind high gauge R&R were available. The QS 9000 MSA manual talks about the precautions to be taken while conducting a gauge R&R study, the possible reasons of high repeatability and reproducibility and various graphical methods of analysing the causes of variation. Combining all these aspects and superimposing our own experience, some general guidelines for investigation of root causes are given below. These guidelines are expected to increase the efficiency of the improvement process and extend gauge R&R studies beyond simple evaluation.

(1) Before suspecting any aspect related to the gauge and appraisers, look back at the way in which the entire exercise has been conducted. Check whether the steps mentioned under the heading 'Preparation for a measurement system study' in the MSA reference manual (chapter II, section 3) have been followed. If you have reason to believe that some of the steps have been ignored, repeat the study incorporating those steps.

(2) If the repeatability error (EV) is high:

(a) Refer back to the range chart constructed with the data. Ensure that points beyond the upper control limit have not been considered for computation of the gauge standard deviation s_0. Either repeat measurements are to be taken for such situations or they have to be discarded before computation of s_0. However, in either situation, the special cause that produced the out of control situation must be identified.

(b) Check whether there is an excessive within-part variation (like taper, ovality, etc.) by studying the profile of the part and collecting appropriate data.

(c) Check whether the gauge is sufficiently rigid.

(d) Check whether the location at which the measurement is to be taken is clearly defined and understood properly by the appraisers. This situation may result in high reproducibility as well.

(e) Check whether the instrument requires maintenance.

(f) Check whether a fixture of some sort is needed to help the appraiser use the gauge more consistently. This may also result in high reproducibility.

(g) If none of the above reasons are found to be valid, brainstorm and find out whether the gauge is suitable for the intended measurement. The suitability can be checked with respect to the type of material (soft/hard/flexible/compressible), the design of the part, the complexity involved in measurement and the environmental conditions. Unsuitable gauges may also lead to high reproducibilty.

(3) If the reproducibility is high:

(a) Plot the appraiser average/run chart in which the averages of the multiple readings by each appraiser on each part are plotted appraiserwise with part number as an index (as shown in Fig. 1). This chart helps in determining the consistency among appraisers. Normalized individuals' chart or Whisker's chart can be used to identify part-appraiser interactions and existence of outliers. If the charts reveal any systematic pattern, identify the reason by consulting the appraisers.

(b) Find out whether all the appraisers are adequately trained in the measurement method.

(c) Check whether the calibrations on the gauge dial are clear.

(d) Check whether the location at which the measurement is to be taken is clearly defined and understood properly by the appraisers. This situation may result in high repeatability as well.

(e) Check whether a fixture of some sort is needed to help the appraiser use the gauge more consistently. This may also result in high repeatability.

(f) If none of the above reasons are found to be valid, brainstorm and find out whether the gauge is suitable for the intended measurement. The suitability can be checked with respect to the type of material (soft/hard/flexible/compressible), the design of the part and the complexity involved in measurement. Unsuitable gauges may also lead to high repeatability.

Reference

Measurement Systems Analysis—Reference Manual, 2nd Edn (Co-published by Chrysler Corporation, Ford Motor Company, General Motors Corporation).