

An overview of capability evaluation of Measurement Systems and Gauge Repeatability and Reproducibility Studies

Ozlem Senvar* and Seniye Umit Oktay Firat

Marmara University, Industrial Engineering Department, Faculty of Engineering, Goztepe Campus, Kadikoy-34720 Istanbul, Turkey

Received: 5 August 2010 / Accepted: 20 August 2010

Abstract. This study is an overview of the capability evaluation of measurement systems. Principally, the determination of capability of a measurement system is an important aspect of quality and process improvement initiatives. In practice, various methods are developed and used for determining measurement system capability. As a measurement tool, potential effectiveness of gauge should be considered and significant factors should be identified for that purpose. A set of procedures referred to as Measurement Systems Capability Studies are conducted for assessing capability of gauge, isolating sources of variability in the system, evaluating how much of total observed variability is due to gauge, and investigating two components of measurement error: repeatability and reproducibility of gauge. Gauge Repeatability & Reproducibility (Gauge R&R) Study tries to estimate repeatability and reproducibility components of measurement system variation with primary objective of assessing whether gauge is suitable for intended application or not. Measurement System Analysis (MSA) is a collection of statistical methods, which includes Gauge R&R Study, for analysis of measurement system capability. In this study, detailed literature review of MSA, Gauge R&R and Measurement Systems Capability Studies, and general discussion of misclassification probabilities that give useful, reliable information about measurement systems performance would be provided.

Keywords: Measurement System Analysis (MSA); Gauge Repeatability and Reproducibility (Gauge R&R); Measurement Systems Capability Studies

1 Introduction

Statistical Process Control (SPC) is a very powerful collection of problem-solving tools that are widely used for monitoring and improving process [1]. The purpose of measurements is to track the process. For this reason, SPC methods use measurement process in order to control the process. Consequently, the quality of the measurement process influences the quality of the process improvement actions. Principally, measurement capability is critical to the quality of measurement.

Since every process produces a product and every product possesses qualities (features) and every quality feature can be measured; there are two types of measurements to verify quality and to quantify performance in manufacturing environment. These are measurement of product and measurement of process. Therefore, product evaluation and process improvement require accurate and precise measurement techniques [2]. Reliability of measurements used in decision making should also be taken into consideration. Thus, entire process of obtaining measurements should involve understanding the impact of

measurement error for decision making about the product as well as process.

Measurement instruments, which are also known as gauges or gages, are used to evaluate the quality of products. Measurement Systems are conceptually much more sophisticated than the measurement instruments used for measuring. The measurement value is a result of the measurement process carried out by a measurement system that is a collection of measuring instruments; operator(s) or appraiser(s); conditions or different points in time under which the instrument is used; environment under which the reading has been obtained; standards; operations; methods used setup and measure the parts; tooling and fixture that locates and orientates the object under measurement; software that performs intermediate calculations and outputs the result; and assumptions used to quantify a unit of measure or the complete process used to obtain measurements. The purpose of a measurement system is to distinguish one part from another.

Industrial measurement systems are being used to measure specific quality attributes of products and their production processes [3]. The effectiveness of a measurement system depends upon accurate gauges and proper gauge use. Measuring equipment and processes must be

* Correspondence: ozlemsenvar@gmail.com

well controlled and suitable to their application in order to assure accurate data collection [4]. Setup or calibration activities and etc. can be the other factors that can influence the measurement system performance [1].

The ideal measurement system must always produce true measurements. However, in industrial practice, this is almost always impossible, that is, measurements contain systematic and random errors. This is why measurement process is subject to all the rules of variation and SPC. In fact, knowledge of variation is one of the most powerful tools for a company to apply in the quest for improvement [5]. Generally, in any activity involving measurements, total variation consists of product variation (variability inherent in the units or items that are being measured) plus measurement variation (variability from the measurement system that is being used). Measurement system variability must be small with respect to product specifications as well as process variation. Understanding the individual components of process variation is critical to the process because the reduction of process variation requires the ability to discriminate between process variation and measurement variation. Apart from this, measurement system must be in state of statistical control. Especially, control charts, which are able to determine whether a process is in a state of statistical control or not, are used for this purpose. Also, statistical methods can be used to separate these components of variance, to access gauge capability. As a matter of fact, SPC is useful in achieving process stability and improving capability through the reduction of variability [1]. Therefore, it is possible to say that quality of the measurement system is characterized by its statistical properties. Also, it should be noticed that as the complexity of the measurement process decreases, the probability of measures that may be affected by errors decreases [6].

Measurement systems capability analysis aims to test if the measurement process is capable, that is, if the variability of a measurement system is small relative to the variability of a monitored process [6]. Measurement capability analyses are critical to the success of every measurement and ensure that future measurements will be representative of the characteristic being measured. Actually, MSA is a very important issue that has to be performed prior to any process improvement initiatives.

Based on brief definitions that are introduced above, this study is carried out as a detailed literature review and a general discussion of gauge and measurement systems capability studies. The outline of the study, in particular order is as follows: background information about Measurement Systems Capability Studies, MSA, Gauge R&R Studies would be examined, respectively. Several ways for summarizing capability of gauge or measurement instrument accompanied by their interpretations as a guideline would be introduced. Discussion of the misclassification probabilities that give useful, reliable information about measurement systems performance would be given, as well.

2 Measurement Systems Capability Studies

Measurement capability is critical to the quality of measurement. Low quality data is not beneficial for a measurement system, which is intended to perform measurement activities [7]. When a measurement system is not capable or suitable, it can affect firstly the related process performance and then business and managerial decisions negatively which can cause a considerable economic loss for an organization. As a matter of fact, with an inadequate measurement system, there would be difficulties in monitoring, controlling, improving and managing a process effectively. A set of procedures, often referred to as Measurement Systems Capability Studies, are widely used to assess the adequacy and quality of measurement systems. The purposes of measurement systems capability studies are to assess capability of a gauge, to isolate sources of variability in the system and to evaluate how much of the total observed variability is due to the gauge [8].

In particular, measurement is the vital part of Six Sigma Methodology. For this reason, MSA is used extensively at the measure stage of DMAIC [9]. MSA is a collection of statistical methods, which includes Gauge R&R Study, for the analysis of measurement system capability [2, 7]. Simply, MSA assesses the adequacy of a measurement system. MSA can also be defined as the way of understanding and managing measurement error since it evaluates as well as tries to minimize the measurement error. Researchers have determined the need for using Gauge R&R Studies as feedback to improve measurement systems [2].

Due to the fact that all measurements contain error, simple but reasonable mathematical model for measurement system capability studies can be developed as follows:

$$y = X + \epsilon \quad (1)$$

where y represents total observed measurement, X stands for true value of measurement on a unit of product, and ϵ is the symbol for measurement error. Notice that, $X \sim N(\mu, \sigma_P^2)$, $\epsilon \sim N(0, \sigma_{\text{Gauge}}^2)$, that is, X and ϵ are normally and independently distributed random variables, with means μ and 0, and variances (σ_P^2) and (σ_{Gauge}^2), respectively.

Variance of total observed measurement (y) is given below:

$$\sigma_{\text{Total}}^2 = \sigma_P^2 + \sigma_{\text{Gauge}}^2 \quad (2)$$

where σ_{Total}^2 is the variance of the measurements, σ_P^2 is the variance of the process and σ_{Gauge}^2 is the variance of the measurement process (variance of the measurement system is broken up into components corresponding to Repeatability and Reproducibility (R&R)) [10].

2.1 Measurement System Analysis (MSA)

When measuring the output from a process, there are two sources of variation that should be considered: part-to-part variation and measurement system variation. Notice

that part refers to the variability of the individual pieces. MSA seeks to identify the variation components in the measurement. MSA evaluates measurement error through the examination of multiple sources of variation in a process, including the variation resulting from the measurement system, from the operators and from the parts themselves [7].

MSA can also be defined as the way of understanding and managing measurement error. It evaluates as well as tries to minimize the measurement error. Actually, MSA is a very important issue that has to be performed prior to any process improvement initiatives. As a matter of fact, measurement error masks true process capability [11]. From this standpoint, MSA has to be performed, before collecting data from the process for analyzing the process capability. To sum up, MSA confirms that the measurement system measures consistently, accurately, and adequately discriminates between parts and is a critical first step that should precede any data-based decision making, including Statistical Process Control (SPC), Correlation and Regression Analysis, and Design of Experiments (DOE).

After emphasizing the definitions of MSA, it is necessary to mention that the total measurement system variation has to be resolved into components or causes of variation. Each of these components has to be isolated and quantified. Only then, reducing the contribution of each one of these error components can be possible. MSA classifies components of measurement system variation into **bias, stability, repeatability, and reproducibility**.

Bias is the statistically significant and systematic error of the measurement result from its true master value, sometimes called the accuracy of a measurement [2]. Accuracy is the ability of the instrument to measure the true value correctly on average [1]. More comprehensive definition of bias can be given as the difference between an observed measurement and a true value or reference value obtained from master or gold standard, or from a different measurement technique known to produce accurate values. Notice that, reference value is also referred to as average accuracy, which is the one-time difference between a measurement result and a known standard [1, 12]. Bias is usually attributed either to an instrument error, that adds (or subtracts) a constant value to each reading. This can be due to a worn out instrument or a parallax like error in the appraiser's evaluating the reading.

Stability or drift is the total variation in the measurement obtained with a measurement obtained on the same master or reference value when measuring the same characteristic, over an extending time period. Stability is also defined as different levels of variability in different operating regimes. Inconsistent operator performance, inadequate standard operating procedure, warm-up effects and environmental factors can lead stability [1]. Stability quantifies a change in bias over time [12].

Stability Study statistically monitors the state of the measurement system over a period of time. A measurement system will induce more variations in the readings due to wear and tear as it gets into use. Each measure-

ment system will go out of stability after different intervals based on their usage. Rather than fixed period schedules, stability test can be an excellent guideline to signal when a measurement system should be taken up for calibration. Using Stability Study to determine the calibration frequency can lead to bottom-line savings for an organization. Stability Test scientifically assures the predictability of the measurement system behaviour over an extended time period.

Repeatability can be defined as reflecting basic inherent precision of the gauge itself where precision is the measure of inherent random variation in measurement system [1]. In its simplest form, repeatability is the measurements variation due to instrument error [12]. In other words, repeatability is "within operator" (one appraiser, one instrument) error, and is usually traced to the gauge itself, and is best considered to be random error [2].

Reproducibility can be defined as the variability resulting from external sources such as operators and their unique techniques, setups, and environmental fluctuations over time [2, 7]. In other words, reproducibility is "between operator" (many appraisers, one instrument) error, and is usually traced to differences among the operators who obtain different measurements while using the same gauge [1, 2].

Table 1 shows methods applicable in the automotive industry for evaluating measurement capability (adopted from [3]). In the table, methods are given in particular order which reflects a practical sequence of their application. The first stage is the incoming inspection of a new system (including its stability determination). The last stage of the assessment is the R&R analysis of the measurement system working in the real production environment.

According to the MSA Reference Manual, MSA defines data quality and error in terms of bias, reproducibility, reliability, and stability [7]. Furthermore, MSA provides procedures to measure each term. Gauge Repeatability and Reproducibility (Gauge R&R) studies has come to incorporate the procedures recommended for measurement of bias, reproducibility, and reliability. Several authors address the use of Gauge R&R Studies for managing of these errors, especially the human aspects of these errors [13].

2.2 Gauge Repeatability & Reproducibility (Gauge R&R Studies)

Gauge R&R Studies are used to act as an audit tool and as a source of feedback to improve the measurement procedure. [14] recommended the analysis of measurements to detect the most important causes for process variation. Total process variation consists of part-to-part variation plus measurement system variation. The purpose of a Gauge R&R Study is to measure how much total process variation is caused by the measurement system.

There are two components of measurement systems variability: Repeatability and Reproducibility.

Repeatability (equipment variation) is the variation due to the measurement instrument or gauge when it is used to measure the same part, repeatedly or several

Table 1. Methods applicable in the automotive industry for evaluating measurement capability, adopted from [3].

Step	Examination	Definition for examination	Applied Statistical Method
1	Stability	Different levels of variability in different operating regimes. Inconsistent operator performance, inadequate standard operating procedure, warm-up effects and environmental factors can lead stability.	Range and average analysis
2	Bias	Systematic error in measurement (difference between observed measurement and reference value).	Range and average analysis
3	Linearity	Differences in observed accuracy and/or precision experienced over range of measurements made by system. Problems with linearity are often the result of calibration and maintenance issues.	Regression analysis
4	Repeatability Reproducibility	Repeatability: within operator variation. Reproducibility: between operator variation.	ANOVA, range and average analysis

times, with the same operator or set-up or in the same time period [8]. In other words, it is the variation observed with one measurement device when used several times by one operator while measuring the identical characteristic on the same part. This can also be thought of as “within operator” variability [2]. To estimate repeatability, each operator measures each part at least twice.

Reproducibility (appraised variation) is the variation arising from different operators, set-ups, or time periods [8]. In other words, it is the variation obtained from different operators using the same measurement device when measuring the identical characteristic on the same part. This can also be thought of as “between operator” variation [2]. To estimate reproducibility, at least two operators must measure the parts. Operators should measure the parts in random order, and the selected parts should represent the possible range of measurements.

When repeatability is large compared with reproducibility, then it is necessary to redesign the gauge and perform proper maintenance of instruments. On the other hand, when reproducibility is large compared with repeatability, then operator needs to be better trained in how to use and read the gauge or instrument. It is the possible case that suitable fixture can be needed to help operators to use gauge more consistently [15].

Total Gauge R&R is the estimate of the combined estimated variation from repeatability and reproducibility [7]. Total variance of the measurements, as it is mentioned previously, is the sum of the variation of total Gauge R&R with part-to-part variation where part is the variability of the individual pieces [2, 7].

Gauge R&R Study tries to estimate the repeatability and reproducibility components of measurement system variation with the primary objective of assessing whether the gauge is suitable for the intended application [7]. Gauge capability study provides reliable estimates of the variation components.

In the analysis of a Gauge R&R Study, Analysis of Variance (ANOVA) approach and Tabular Algorithm are used. ANOVA approach is followed by estimation of the

appropriate variance components. Tabular algorithm relies on the range method to estimate the standard deviations of the components of gauge variability. ANOVA approach is widely used by practitioners for analyzing the results of experiments as it is easy to be performed and it can be adapted to deal with very complex experiments. Additionally, it allows obtaining efficient point estimation as well as confidence interval estimation of the variance components associated with the sources of variability in the experiments and it is fairly easy to understand the properties of these confidence intervals. On the other hand, the tabular approach cannot be applied to any study other than the traditional two-factor design, and it does not allow obtaining confidence intervals [8]. For more details about the comparisons of ANOVA approach and Tabular Algorithm, interested readers are referred to [9].

There is another point that has to be emphasized that is, criteria used to access capability of a measurement system, called measurement capability metrics [10]. Several ways that summarize capability of gauge or measurement instrument are listed below:

1. R&R criterion
2. Precision-to-tolerance ratio (P/T)
3. Ratio of process (part) variability to total variability (ρP)
4. Ratio of measurement system variability to total variability (ρM)
5. Signal-to-noise ratio (SNR)
6. Discrimination ratio (DR)

Before explaining these ways, the parameters must be mentioned because they describe the variation in the monitored process and the variation in the measurement system, in a Gauge R&R Study. These parameters are given in Table 2.

When measurement system variability is less than 10% of total variability, measurement system is considered acceptable. When measurement system variability is between 10 and 30%, measurement system can be

Table 2. Parameters, which describe the variation in the monitored process and the variation in the measurement system, in a Gauge R&R Study.

Parameter	Definition of variation
γP	Symbolizes variance of the monitored process
γM	Symbolizes variance of the measurement system
$\gamma T = \gamma P + \gamma M$	Formulates total variance of the response variable
$\rho P = \gamma P / \gamma T$	Formulates proportion of total variance due to process
$\rho M = \gamma M / \gamma T$	Formulates proportion of total variance due to measurement system

acceptable. When measurement system is above 30% variability, measurement system cannot be considered acceptable [15].

2.2.1 %R&R criterion

From the Gauge R&R Study, the estimated variance components are obtained. One basic criterion for the acceptability of the measurement system is commonly referred to as the Total gauge R&R %Study Variance that should be suitably small. %Study Variance refers to the per cent study variation which involves ratios of standard deviations [10]. According to AIAG [16], values less than 0.1 indicate that measurement system is considered to be acceptable. Values between 0.1 and 0.3 indicate that measurement system may be acceptable depending on factors such as the importance of the application, the cost of the measurement device, and the cost of repair. Values over 30% indicate that measurement system is generally considered unacceptable and every effort should be made to improve the measurement system.

2.2.2 Precision-to-Tolerance (P/T) ratio

There is a common practice for comparing the estimate of gauge capability with respect to width of specifications or tolerance band ($USL - LSL$) for part that is being measured. This ratio is called Precision-to-Tolerance (P/T) ratio, which is a function of γM . P/T is expressed as follows:

$$P/T = (k \sqrt{\gamma M}) / (USL - LSL) \times 100\% \tag{3}$$

where USL and LSL are upper and lower specification limits, respectively. Popular choices for constant k , $k = 5.15$ and $k = 6$. Value $k = 5.15$ corresponds to limiting value of number of standard deviations between bounds of 95% tolerance interval that contains at least 99% of normal population. Value $k = 6$ corresponds to number of standard deviations between usual natural tolerance limits of normal process.

Values of estimated Precision-to-Tolerance ratio (P/T) of 0.1 or less indicate that measurement system is adequate. This is based on generally used rule that requires measurement device to be calibrated in units one-tenth as large as accuracy required in final measurement [1].

P/T is not capable to give a good indication of how well a measurement system performs for a particular process because a process with a high capability can tolerate a measurement system with a higher P/T than a process that is not as capable [17]. For this reason, the adequacy of a process can also be determined by other measures of measurement system capability. One of these is function of ρP , that is the ratio of process (part) variability to total variability. Another is function of ρM , that is the ratio of measurement system variability to total variability, since $\rho M = 1 - \rho P$.

2.2.3 Signal-to-Noise Ratio (SNR)

Another measure of measurement system adequacy is Signal-to-Noise Ratio (SNR), which can be written as a function of ρP in the following way:

$$SNR = \sqrt{\frac{2\rho P}{1 - \rho P}} \tag{4}$$

SNR is the number of distinct levels or categories that can be reliably obtained from the measurements. A value of 5 or greater is recommended and a value of less than 2 implies inadequate gauge capability.

The number of distinct categories (ndc) that a measurement system can identify is used as a capability metric. There are several metrics that report ndc . The ndc has also been reported as the Signal-to-Noise Ratio (SNR). For more details on ndc , interested readers are referred to [10].

2.2.4 Discrimination Ratio (DR)

Another measure of gauge capability is Discrimination Ratio (DR), which can be expressed as another function of ρP in the following way:

$$DR = \frac{1 + \rho P}{1 - \rho P} \tag{5}$$

Discrimination Ratio (DR) can be used to determine the number of categories that the measurement system is capable of distinguishing [10]. The number of data categories is often referred to as the discrimination ratio since it describes how many classifications can be reliably distinguished given the observed process variation. DR must

exceed 4. This indicates for an adequate gauge capability [7]. For more details on the measurement capability metrics, interested readers are referred to [1, 8, 10].

It has to be taken into account that none of these metrics really describe gauge or measurement system capability in any sense that is directly interpretable. The metrics mentioned previously have been criticized for being too subjective. Unlike the measures ndc , SNR , and DR , misclassification rates provide a more objective measure of the system capability. Unlike the number of categories metrics, the misclassification rates are dependent upon the specification limits.

2.3 Misclassification rates: false defectives and passed defectives

A less subjective measure of an adequate measurement system discussed in the literature is the misclassification rate a.k.a. misclassification probability, since it is based on the actual performance of the measurement process. For this reason, misclassification rates are accepted as more objective measures of capability. In other words, misclassification rates provide a more objective measure of the gauge or measurement system capability [10].

Effective capability of measurement system is best described in terms of how well it discriminates between good and bad parts. From the results of a standard measurement systems capability study, calculation or computation of misclassification probabilities can be easily performed. These probabilities provide reliable, useful, easy-to-understand information about measurement systems performance [1].

The adequacy of a measurement system can be determined by its ability to distinguish between good and bad parts. Misclassification rates are explained in [10] by considering the following model: $Y = X + \epsilon$, where Y is the measured value of a randomly selected part, X is the true measurement of the part and ϵ is the measurement error. X and ϵ are assumed to be normally and independently distributed random variables, that is, $X \sim N(\mu_P, \sigma_P^2)$ and $\epsilon \sim N(\mu_M, \sigma_\epsilon^2)$. Y is normally distributed with mean $\mu_Y = \mu_P + \mu_M$ and variance $\sigma_Y^2 = \sigma_P^2 + \sigma_\epsilon^2$. Mean μ_M is assumed to be 0 since it is accepted as measurement bias.

The two following cases are also considered:

Case 1: A unit of product or part is in conformance to the specifications, if $LSL < X < USL$.

Case 2: Measurement system will pass a product or part as a non-defective, if $LSL < Y < USL$.

There are two types of possible misclassifications of a product or part:

1. If Case 1 is true but case 2 is false, that is, a conforming product or part has been incorrectly failed. This is called a false failure, which can also be defined as the misclassification of a good product or part as bad. This occurs when the product or part really is in conformance, but is not passed.
2. If Case 1 is false but case 2 is true, that is, a nonconforming product or part has been incorrectly passed.

This is called a missed fault, which can also be defined as the misclassification of a bad product or part as good. This occurs when the product or part is not in conformance, but the measurement system passes the product or part.

It should be taken into account that both types of misclassifications can be quite costly.

Furthermore, the probability of a false failure is sometimes referred to as producer's risk whereas the probability of a missed fault is sometimes referred to as consumer's risk. It is a very useful way to determine capability of measurement system in terms of producer's risk and consumer's risk. Producer's risk is defined as joint probability or conditional probability that a measurement system will fail a product or part when the product or part conforms to specifications. Consumer's risk is defined as joint probability or conditional probability that a measurement system will pass a product or part when the part does not conform to specifications. Again, risks of producer and consumer can be quite costly. Notice that, the probabilities of these types of misclassifications can be calculated using either joint probabilities or conditional probabilities. The joint probabilities are illustrated in this study. For computation of conditional probabilities, see [1].

The probability of a false failure is as follow:

$$\delta = P(LSL < X < USL \text{ and } (Y < LSL \text{ or } Y > USL)). \quad (6)$$

The probability of a missed fault is as follow:

$$\beta = P((X < LSL \text{ or } X > USL) \text{ and } (LSL < Y < USL)). \quad (7)$$

If δ , β , or both are unacceptably large, then the measurement system is not acceptable. Unfortunately, there are no standard acceptable levels to which these values can be compared.

The likelihood or probability that a system will misclassify parts is an important metric that should be estimated in addition to the metrics corresponding to the other criteria. However, there are no single point estimates available for the misclassification rates, only lower and upper bounds for a given level of confidence [10].

In practice, true values of parameters μ , σ_P^2 , σ_T^2 are generally unknown. If one uses only point estimates, calculation does not account for uncertainty in the estimates. It would be very helpful to provide confidence intervals for these parameters in calculation of δ and β . One way to do this is, to compute δ and β under different scenarios (pessimistic scenario and optimistic scenario) suggested by confidence intervals on variance components [1]. One might consider a pessimistic scenario with the worst possible performance for the measurement system combined with the worst process capability for the manufacturing process. In other words, pessimistic scenario is computed assuming the worst possible performance for measurement system as well as manufacturing process. To do this, set σ_P^2 equal to upper bound of confidence interval for σ_P^2 and solve for value of σ_{Total}^2 that provides lower bound on ρP . This is done by computing δ and β using upper bound on

σ_P^2 and lower bound on ρP . Conversely, one might consider an optimistic scenario with the best possible performance for the measurement system combined with the best process capability for the manufacturing process. In other words, optimistic scenario is computed assuming the best possible performance for measurement system as well as manufacturing process. For more details on the topic, interested readers are referred to [1].

The confidence intervals are quite wide, as a result of the small number of operators suggested by AIAG [7]. Researchers have shown and proven that when the operator effect is considered random, the number of operators in a Gauge R&R Study should be larger than the standard use of three. Therefore, it is recommended to increase the number of operators in any Gauge R&R Study, when operator effect is random.

Zappa and Deldossi (2009) showed how to exploit the misclassification rates to set the critical levels of measurement capability indices and to find the number of parts, operators and replications controlling for the conditional probability errors [6].

3 Conclusions

We have mentioned Measurement Systems Capability Studies, Measurement System Analysis (MSA) and Gauge Repeatability and Reproducibility (Gauge R&R) Studies based on empirical and theoretical surveys.

Based on our review of empirical studies, we realized that a gap appears between the actual use of measurements by manufacturing professionals and the theories of gauge control in measurements. Most of the empirical surveys seek to explore that theory-practice gap, specifically focusing on the use of Gauge Repeatability and Reproducibility (Gauge R&R) Studies, in both physical measurements and visual inspection processes for the manufacturing environment. Developing and establishing standards for visual inspection processes and developing accepted methods for improvement of inspection accuracy should be a top priority for practitioners. Generally, empirical surveys highlight the practicality of the Gauge R&R Study as a decision-making tool for practitioners.

Based on theoretical surveys, point estimates alone do not completely describe gauge capability. Similarly, point estimates of a process capability ratio do not completely describe process capability. In both situations, confidence intervals are of considerable practical value and should become part of any standard Gauge R&R Study. There are several ways, measurement capability metrics that summarize capability of gauge. Much of the guidance regarding interpretation of these metrics is arbitrary and cannot provide direct information on the capability of the gauge in order to discriminate between good and bad parts. For this reason, it might be beneficial to consider the confidence intervals for δ and β in the decision process.

Most of the surveys focus on the methods that are based on the assumption that the response variable from the experiment is normally distributed. However, in industrial practice, there are situations where measurement error is not normally distributed. Methodology for this issue needs to be developed and carried into real life applications.

References

1. D.C. Montgomery, *Introduction to Statistical Quality Control*, 6th edn. (John Wiley & Sons, New York, 2009)
2. Rathel R. Dick Smith, S.W. McCrary, R. Neal Callahan, Gauge repeatability and reproducibility studies and measurement system analysis: A multimethod exploration of the state of practice, *J. Ind. Technol.* **23**, 1-12 (2007)
3. R. Tabisz, The capability evaluating of industrial measurement systems, in *XVII IMEKO World Congress June 22-27, Dubrovnik, Croatia, 2003*
4. T. Little, 10 Requirements for Effective Process Control: A Case Study, *Qual. Progress* **34**, 46-52 (2001)
5. B.L. Joiner, M.A. Gaudard, Variation, management, and W. Edwards Deming, *Qual. Progress* **23**, 29-36 (1990)
6. D. Zappa, L. Deldossi, Misclassification rates, critical values and size of the design in measurement systems capability studies, *Appl. Stoch. Mod. Bus. Ind.* **25**, 601-611 (2009)
7. Automotive Industry Action Group (AIAG), *Measurement systems analysis*, 3rd edn. (Author, Southfield, MI, 2002)
8. R.K. Burdick, C.M. Borror, D.C. Montgomery, A review of methods for measurement systems capability analysis, *J. Qual. Technol.* **35**, 342-354 (2003)
9. D.H. Ballard, D.W. McCormack Jr., T.L. Moore, J. Prins, P.A. Tobias, M. Pore, A Comparison of Gauge Study Practices, in *ASA Proc. of the Section on Quality and Productivity, 1997*, pp. 31-36
10. W.H. Woodall, C.M. Borror, Some relationships between Gage R&R Criteria, *Qual. Reliab. Eng. Int.* **24**, 99-106 (2008)
11. M.J. Harry, J.R. Lawson, *Six sigma producibility analysis and process characterization* (Addison-Wesley, New York, 1992)
12. D.W. Hoffa, C. Laux, Gauge R&R: An effective methodology for determining the adequacy of a new measurement system for micron-level metrology, *J. Ind. Technol.* **23**, 1-9 (2007)
13. D.H. Besterfield, *Quality Control*, 7th edn. (Prentice Hall, Englewood Cliffs, New Jersey, 2004)
14. R. Lupan, I.C. Bacivarof, A Relationship between Six Sigma and ISO 9000:2000, *Qual. Eng.* **17**, 719-725 (2005)
15. J. Antony, M. Kumar, M.K. Tiwari, An application of six sigma methodology to reduce the engine-overheating problem in an automotive company, *J. Eng. Manuf. B8* **14**, 633-646 (2005)
16. Automotive Industry Action Group (AIAG), *Statistical Process Control*, 2nd edn. (Author, Southfield, MI, 2005)
17. D.P. Mader, J. Prins, R.E. Lampe, The Economic Impact of Measurement Error, *Qual. Eng.* **11**, 563-574 (1999)