Role of Measurement System Analysis in Capstone Production

Mustafa Shraim
Ohio University
shraim@ohio.edu

William Neil Littell
Ohio University
littellw@ohio.edu

Zaki Kuruppalil
Ohio University
kuruppal@ohio.edu

Abstract

A senior capstone course at the college of engineering of a major Midwestern university will integrate all of its requirements by using the production part approval process (PPAP). The PPAP is a method that component suppliers of the automotive manufacturers use to properly carry out and document all customer requirements. The PPAP methodology was developed by the Automotive Industry Action Group (AIAG) and primarily used in the supply chain to establish confidence in component suppliers and their production processes. It includes conducting several studies and analyses such as failure mode and effects analysis (FMEA), advanced product quality planning (APQP), and measurement systems analysis (MSA), among others. This paper will focus on the MSA component and the validation of measurements using gauge repeatability and reproducibility (GR&R) methods. The aim of this paper is to introduce GR&R into the senior capstone process by answering two questions: Are students aware of how significant measurement system variation can be when compared to total variation? What can be done if the GR&R study results do not satisfy standard requirements? An example with capstone production data is presented.

Introduction

The increase in outsourcing of components and subassemblies to external suppliers, including those offshores, created the need for standardizing the approval process. The PPAP is a valuable and rigorous program for establishing confidence in the manufacturing process of component and subassembly suppliers (AIAG Work Group, 2006). PPAP initially started in the automotive industry but was adopted by other industries over time. It is becoming a common language for communicating expectations to suppliers regarding the qualification of the manufacturing process. Therefore, it is imperative that engineers entering the workforce have a hands-on understanding of PPAP and the interactions among its parts.

The PPAP manual, along with other relevant documents, is published by the Automotive Industry Action Group (AIAG). This non-profit organization was initially founded in 1982 by representatives of the "Big 3" automotive manufacturers in North America: Ford, General

Motors, and Chrysler. Since then, other manufacturers, both in the automotive as well as other industries, have become members of the organization. This includes, but not limited to, original equipment manufacturers and their Tier 1 suppliers. In the automotive industry, the quality management system requirements are governed by the International Automotive Task Force through its standard, IAFF 16949 (2016). This standard includes all requirements in ISO 9001 (2015) as its core in addition to other requirements, such as production part approval.

Generally, there are five submission levels under PPAP. These levels determine what is involved, the documentation required, as well as sample submission requirements. Table 1 below shows the levels as published in the PPAP manual by AIAG.

Table 1. PPAP submission levels.

Level	Description of Requirements
1	Submission warrant and designated appearance approval report
2	Submission warrant, with product samples and some supporting data required by the customer
3	Submission warrant, product samples, and complete supporting documents
4	Submission warrant and other requirements as defined by the customer
5	Submission warrant, product samples, and complete supporting documents at the supplier's manufacturing site

Typically, level 1 is applicable when minor changes of an already-approved PPAP submission occur. For example, physically moving the manufacturing process within a facility may trigger a level 1 submission. On the other hand, level 5 is for initial submissions of parts that involve safety features, while level 3 is the default submission for most situations.

Depending on the level, each submission consists of up to 18 items that must be either submitted and/or retained at the manufacturing location for review by the customer upon request. Examples of items submitted are design FMEA, design flow diagram, process FMEA, control plan, and measurement system analysis (MSA), among others. The focus of this paper will be on MSA with GR&R studies as its analysis output (AIAG Work Group, 2010).

Measurement Systems

A measurement system is the process used to acquire data for a quality characteristic of a given product so that a decision can be made on its status. Ideally, no error comes from the measurement system. In practice, however, such an objective is impossible to realize due to the variation in the measuring equipment as well as between and within operators performing the measurement (appraisers). When an error exists in the measurement system, the true value of the characteristic being measured could be either overestimated or underestimated.

If this deviation is substantial relevant to the tolerance, it may lead to one of the two types of errors or mistakes (Deming, 1982; Montgomery, 1991):

1. *Type I Error*: This type occurs when good product that genuinely conforms to established specifications is deemed unacceptable. This is also referred to as “false alarm” or “producer’s risk.” This is equivalent to treating variation coming from common cause as if it is a particular cause.
2. *Type II Error*: This type results when the nonconforming product is deemed acceptable and moved on to the next stage (e.g., customer). This is often referred to as “consumer’s risk.” This is equivalent to not reacting to special-cause variation.

The difference between the true value and the measured value can be due to accuracy, precision, or both. The accuracy is the deviation of an average of repeated mesurements from the true value while precision refers to the scatter of measurement around the true value (Juran & Gryna, 1980). A rule of thumb is that the precision of the measuring equipment should be such that its total variability is one-tenth of the the tolerance being measured (Feigenbaum, 1991). For example, for our part of tolerance of 0.0300 inches, the variability from the measuring equipment should not exceed 0.0030. This equipment should be capable of reading 0.003-inch calibration marks.

These two types of errors may occur when the measurement system is inadequate. That is, the variation attributed to its components is too large when compared to total variation. Figure 1 shows the breakdown of sources of variation.

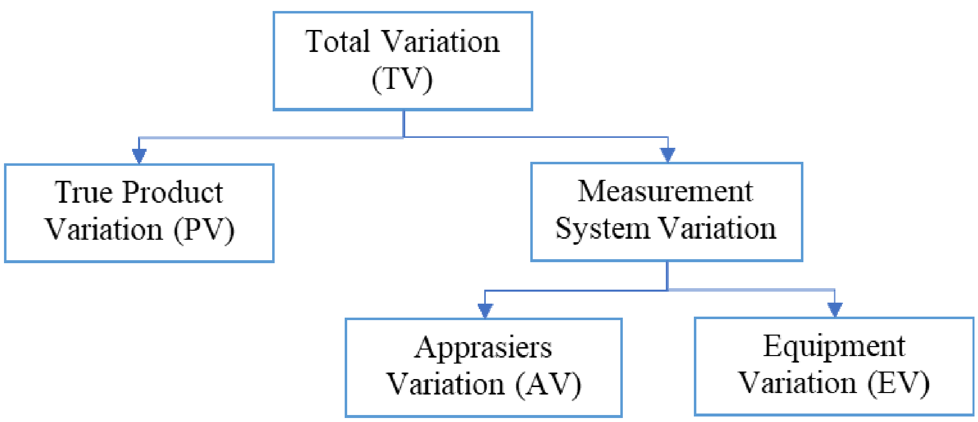

Figure 1. Breakdown of variation components.

As shown in Figure 1, TV in a measured characteristic is comprised of the actual or part-to-part variation as well as the measurement system variation (MSV) (Barrentine, 1991). If the measurement system includes appraisers in addition to the measuring equipment, then MSV must account for both. Mathematically, Figure 1 above can be represented by the following equations using the variance components:

$$\sigma^2_{Total} = \sigma^2_P + \sigma^2_{MS}$$
$$\sigma^2_{MS} = \sigma^2_{Appraisers} + \sigma^2_{Equipment} \quad (1)$$

The objective of MSA is to quantify measurement errors by assessing the two sources of variation (appraisers and measuring equipment). Variability attributed to the appraisers is commonly referred to as reproducibility, and the one attributed to the measuring equipment is referred to as repeatability (AIAG-Work Group, 2010). Thus, variation (standard deviation) for y (GR&R) can be expressed as follows:

$$\sigma_{GRR} = \sqrt{\sigma^2_{Appraisers} + \sigma^2_{Equipment}} \quad (2)$$

In analyzing measurement systems, it is often desirable to break down the GR&R variation into its components. By separating these sources of variation in the measurement system, effective countermeasures can be applied. For example, if GR&R variation is mostly due to the measuring equipment, it may be more useful to repair, calibrate, or replace. On the other hand, if appraisers show they are not consistent (or their results are not reproducible), it may be time to invest in standardizing methods of measurement as well as training of appraisers.

There are different methods used in analyzing measurement system or gauge studies (Wheeler, 2013). Individually, these methods are

- Analysis of variance (ANOVA) (Burdick, Borror, & Montgomery, 2003)
- AIAG (X-bar and R) (AIAG, 2006)
- Evaluating the measurement process (EMP) method (Wheeler & Lyday, 1989)

These methods differ in the way they estimate the component standard deviations before the overall GR&R variation and contribution are calculated. These analyses are included in many of the commonly available statistical analysis software available. In this analysis, we utilize the AIAG method but will also run the analysis using the ANOVA method for comparison.

In the automotive industry, it is a standard practice to require the GR&R variation to be within 10% of total variation. If the GR&R variation is greater than 30% of total variation, then the measurement system is considered unacceptable. For the situations where the GR&R is between 10% and 30%, the results could be acceptable or conditionally acceptable, depending on the application, among other factors (AIAG, 2006). When the measurement system is deemed unacceptable, it is customary that corrective action is applied and the GR&R study is run again.

As mentioned previously, this paper aims to introduce GR&R into the senior capstone process by answering two questions: Are students aware of how significant measurement system variation can be when compared to total variation? What can be done if the GR&R results do not satisfy standard requirements? For the first question, a survey was conducted for the overall PPAP perception with multiple open-ended questions. Among the questions or requested information, students were asked to describe the role of variation in both the part and the process. For the second question, GR&R studies were conducted and analyzed.

Results and Discussion

Among the questions or requested information in the survey, students were asked to describe the role of variation in both the part and the process. From the 25 students in the class, almost all the responses were addressing variation in the manufacture of the parts and not necessarily the measurement system. Here are some examples of their responses:

- “Variation in the process creates variation in the parts produced”
- “Variation in the part is the difference between parts that are theoretically supposed to be the same. This type of variation is extremely common and does not necessarily create quality issues. Variation in the process is variation in how the part is made, and this is a much more destructive type of variation. This commonly causes quality issues and creates many problems in manufacturing”
- “No matter how precise a part or process is there will always be variation. it is impossible to produce parts consistently that are perfectly the same. with that variation in parts is how parts differ from one another. variation in process is the difference in the same process by different suppliers that yield different results”
- “It is desirable to have as little variation as possible. This is because it will help reduce product defects and ensure that there is a consistency with the parts being produced”
- “Variation in the part is due to variation in the process. If the company making the part has a process that is not tightly controlled, there will be variation in the part”

A couple of responses indirectly addressed the issue of the measurement systems:

- “It is important to get your process under control then you can start improving it”
- “Processes have variation such as operator methods and external factors that need to be controlled”

The above results clearly show that students do not think of measurement system variation when they they are asked about part and process variation. Instead, they only think of variation in making the product. However, and as shown in Figure 1, overall part variation (perception) is made up of two components: part-to-part (true variation) and measurement system (error).

To answer the second question, three teams of the senior capstone class participated in this GR&R study with three appraisers from each team. These are the same teams that will eventually complete the PPAP process in the second capstone class. Before the start of the study, an overview of measurement system analysis was conducted to explain how measurement system variation is related to overall variation. Upon the completion of the overview, three volunteers from each team completed the GR&R studies in two phases:

- Phase I: Conduct the GR&R study per instructions given
- Phase II: Conduct the GR&R study by making simple improvements to minimize variation explained in the overview. The measuring method may vary from team to team.

A set of 10 parts similar to components that will potentially be in a capstone project were used to measure a designated dimension by three appraisers from each team using a provided pair of calipers. Each team used the same pair of calipers in both phases. In addition to the three appraisers, the study was coordinated by faculty members to ensure integrity of results. Table 2 outlines the process for completing the measuring and recording process. In this study, three trials were used.

Table 1. Gauge study instructions.

Step	Instruction
1	Appraiser 1 measures all 10 samples in a random order
2	After an appraiser reads each measurement, the coordinator verifies and records it
3	Steps 1 and 2 are repeated for the other appraisers
4	For each additional trial, have each of the appraisers repeat steps 1 and 2

For each phase, 90 measurements per team were taken (10 parts x 3 appraisers x 3 trials). Statistical software was used to analyze the data using the AIAG method for both phases (before and after improving or standardizing the measurement process by each team). It should be mentioned that 6 standard deviations were used, as the overall process spread covering 99.73% of the area under the normal distribution. The default for the AIAG is to use 99% of the area which spreads over 5.15 standard deviations. Additionally, the ANOVA method was run once, along with the AIAG method, for comparison purposes. Tables 3, 4, and 5 display the results for the three teams in Phase I.

Table 2. Team 1 results before standardization.

Appraisers: Team 1			Specifications: 0.3850 ±0.0150	
Source	Std Dev.	6 x Std. Dev.	% Study Var	% Tolerance
Total GR&R	0.0028227	0.0169362	59.89	56.45
Repeatability	0.0023436	0.0140615	49.72	46.87
Reproducibility	0.0015733	0.0094398	33.38	31.47
Part-to-Part	0.0037747	0.0226483	80.08	75.49
Total Variation	0.0047134	0.0282803	100.00	94.27

Table 3. Team 2 results before standardization.

Appraisers: Team 2			Specifications: 0.3850 ±0.0150	
Source	Std Dev.	6 x Std. Dev.	% Study Var	% Tolerance
Total GR&R	0.0027283	0.0163697	63.69	54.57
Repeatability	0.0023633	0.0141796	55.17	47.27
Reproducibility	0.0013633	0.0081795	31.82	27.27

Part-to-Part	0.0033029	0.0198172	77.10	66.06
Total Variation	0.0042840	0.0257039	100.00	85.68

Table 4. Team 3 results before standardization.

Appraisers: Team 3			Specifications: 0.3850 ±0.0150	
Source	Std Dev.	6 x Std. Dev.	% Study Var	% Tolerance
Total GR&R	0.0024537	0.0147219	65.49	49.07
Repeatability	0.0021171	0.0127026	56.51	42.34
Reproducibility	0.0012403	0.0074417	33.11	24.81
Part-to-Part	0.0028310	0.0169862	75.57	56.62
Total Variation	0.0037464	0.0224781	100.00	74.93

The commonly reported results for GR&R studies are under the "% Study Var" column. This indicates the extent of GR&R variation when compared to the overall variation. The average GR&R variation for the three teams for Phase I is about 63%. In other words, 63% of the total variation is due to the measurement system, which is not acceptable according to the AIAG standards (2006). Furthermore, on average, there was more variation in the measuring equipment (repeatability) when compared to appraisers (reproducibility), of approximately 54% to 33%. This is an indication that the measuring equipment needs attention such as maintenance, calibration, or possible replacement. This was not within the scope of the study. The percentages in GR&R and part-to-part components do not add to 100% since the standard deviations cannot be added but rather the variances as shown in the previous section.

The "% Tolerance column" in Tables 3, 4, and 5 refers to the percentage of component variation as compared to specifications. For example, GR&R component in Table 5 shows 49.07% contribution of the "6 x Std Dev" value of 0.0147 inches when compared against the tolerance of 0.0300 inches. This information could be relevant if the overall process capability is high that measurement error may be deemed insignificant.

For Phase II, the teams were asked to come up with ways to standardize the method of measuring the parts. Without further involvement from faculty, each team devised simple steps to standardize how they measure the parts to improve consistency between appraisers. Below are some examples of standardization the teams used:

- Orienting part with respect to a marking on the part
- Ensuring perpendicularity between part and calipers; standing the part on end on the table, placing the calipers on the table, and using the center of the calipers
- Placing the sample completely in the calipers to maximize contact area across the surface of the device

The teams repeated the GR&R study for Phase II and data were analyzed in a similar manner. Results are displayed in Tables 6, 7, and 8 below.

Table 6. Team 1 results after standardization.

Appraisers: Team 1			Specifications: 0.3850 ±0.0150	
Source	Std Dev.	6 x Std. Dev.	% Study Var	% Tolerance
Total GR&R	0.002687	0.016123	49.21	53.74
Repeatability	0.001142	0.006854	20.92	22.84
Reproducibility	0.002432	0.014594	44.55	48.65
Part-to-Part	0.004753	0.028520	87.05	95.07
Total Variation	0.005460	0.032762	100.00	109.21

Table 7. Team 2 results after standardization.

Appraisers: Team 2			Specifications: 0.3850 ±0.0150	
Source	Std Dev.	6 x Std. Dev.	% Study Var	% Tolerance
Total GR&R	0.0011903	0.007142	21.07	23.81
Repeatability	0.0009256	0.0055537	16.39	18.51
Reproducibility	0.0007484	0.0044905	13.25	14.97
Part-to-Part	0.0055223	0.0331336	97.75	110.45
Total Variation	0.0056491	0.0338946	100.00	112.98

Table 8. Team 3 results after standardization.

Appraisers: Team 3			Specifications: 0.3850 ±0.0150	
Source	Std Dev.	6 x Std. Dev.	% Study Var	% Tolerance
Total GR&R	0.002468	0.014806	54.19	49.35
Repeatability	0.002058	0.012348	45.19	41.16
Reproducibility	0.001362	0.008169	29.90	27.23
Part-to-Part	0.003827	0.022963	84.04	76.54
Total Variation	0.004554	0.027322	100.00	91.07

The results in Tables 6, 7, and 8 show that, on average, GR&R variation was reduced from 63% in Phase I to 41.5% in Phase II or about one-third. Although this is still not acceptable according to the AIAG standards, it is a significant improvement made only by introducing simple steps to standardize the process. Figure 2 compares GR&R results between the two phases for the three teams.

As mentioned in Phase I "% Tolerance column" in Tables 3 through 8 is the percentage of component variation as compared to specifications. Table 8 shows a GR&R component with 49.35% contribution ("6 x Std Dev" value of 0.0148 inches when compared against the tolerance of 0.0300 inches).

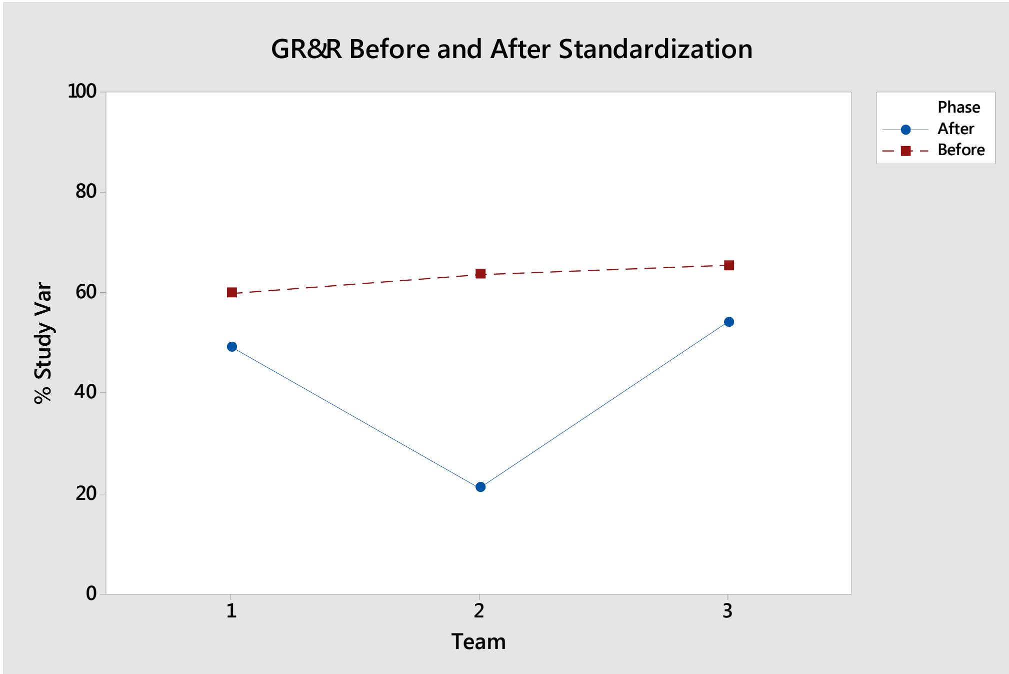

Figure 2. GR&R result.

The ANOVA method for analyzing the GR&R data was also run for one case to compare against the AIAG (or X-bar and R) method. The results are shown in Tables 9, 10, and 11.

Table 9. Two way ANOVA with interaction (Team 2, Phase II).

Source	DF	SS	MS	F	P
Parts	9	0.0027358	0.000304	203.655	0.000
Appraisers	2	0.0000387	0.0000193	12.960	0.000
Parts * Appraisers	18	0.0000269	0.0000015	1.311	0.214
Repeatability	60	0.0000683	0.0000011		
Total	89	0.0028697			

Table 9 shows that the interaction between parts and appraisers is not significant. This means that the variation there is considered random. As a result, this component can be combined to the "repeatability" component as shown in Table 10. In Figure 3, the chart on the bottom right shows close to parallel lines for operators across parts indicating no significant effect for the interactions between parts and appraisers.

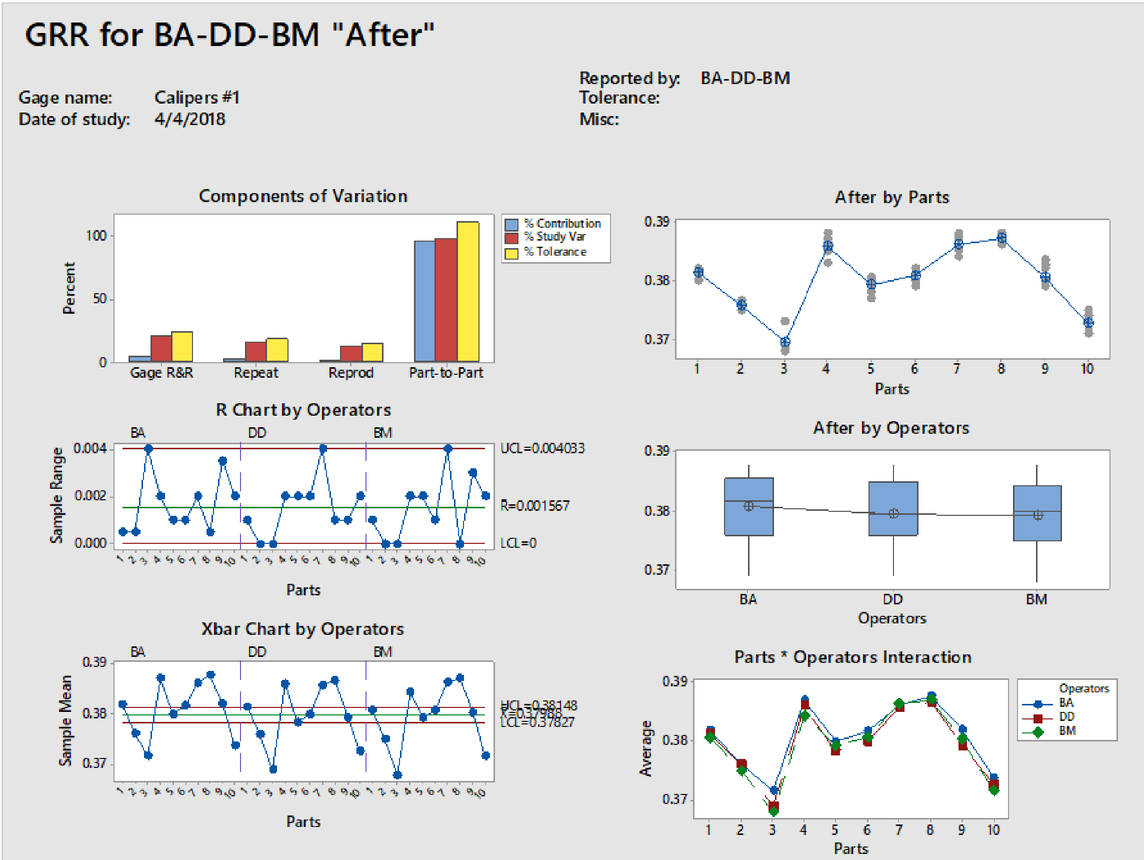

Figure 3. Example of GR&R results (Team 2, Phase II).

Table 10. Two way ANOVA without interaction (Team 2, Phase II).

Source	DF	SS	MS	F	P
Parts	9	0.0027358	0.000304	203.655	0.000
Appraisers	2	0.0000387	0.0000193	12.960	0.000
Repeatability	78	0.0000952	0.0000012		
Total	89	0.0028697			

The GR&R and component contributions based on the ANOVA method are calculated and compared to the AIAG method in Table 11 (ISO, 2015; AIAG Work Group, 2010). These results show a slight difference between the two methods. However, ANOVA determines if there is an interaction effect between parts and appraisers. If the interaction is significant, then this should be investigated and corrected before running the study again.

Table 11. GR&R results using AIAG vs. ANOVA.

Source	AIAG % Study Var	ANOVA % Study Var
Total GR&R	21.07	22.68
Repeatability	16.39	18.55
Reproducibility	13.25	13.05
Part-to-Part	97.75	97.39
Total Variation	100.00	100.00

Concluding Remarks

As PPAP methods are being introduced to the senior capstone classes, students need to understand the importance of the measurement system process and analysis. This paper demonstrated through a survey that students were not aware of the relationship between measurement system variation and total variation. In addition, measurement system studies were conducted to show, for example, that measurement variation can inflate the total variation, which could result in making erroneous decisions. These decisions can have an impact on the bottom line through increasing failure costs of quality. They can also have a detrimental effect on customer satisfaction as nonconforming product may be received by the customer due to committing errors of Type II.

With an average GR&R variation of 63% of total variation, the measurement system analysis in Phase I showed how large the measurement system error could be when compared to total variation. In Phase II, by applying simple steps to standardize the measurement process among appraisers, the GR&R percent contribution was reduced by one-third. These results will be shared with the teams involved at the beginning of their second capstone class.
As shown in the results of Phase II (Tables 6, 7, and 8), Team 2 appeared to reduce GR&R variation by two thirds–from 63.69% to 21.07%. It would be worthwhile to see if the other two teams could realize similar results should they follow the measurement process employed by Team 2. Furthermore, the measuring equipment variation could be reduced by investigating the device for calibration status or using an alternate. Finally, another study could be conducted using a high-precision, no-contact measuring system to estimate overall variation then compare data to results in this study.

References

AIAG-Work Group. (2006). *Production part approval process (PPAP), reference manual*. (4th ed.). Southfield, MI: AIAG.

AIAG-Work Group. (2010). *Measurement systems analysis, reference manual*. (4th ed.). Southfield, MI: AIAG.

Barrentine, L. (1991). *Concept for R&R studies*. Milwaukee, WI: ASQC Quality Press.

Burdick, R., Borror, C., & Montgomery, D. (2003). A review of methods for measurement systems capability analysis. *Journal of Quality Technology*, *35*(4), 342-354.

Deming, W. E. (1982). *Out of the crisis*. Cambridge, MA: MIT Press.

Feigenbaum, A. V. (1991). *Total quality control*. (3rd ed.). New York: McGraw-Hill.
International Automotive Task Force. (2016). *Automotive quality management system standard* (4th ed.). Southfield, MI: AIAG.
ISO. (2015). *ISO 9001:2015 quality management systems – Requirements*. Geneva, Switzerland: International Organization for Standardization.
Juran, J., & Gryna, F. (1980). *Quality planning and analysis*. (2nd ed.). New York: McGraw-Hill.
Montgomery, D. (1991). *Introduction to statistical quality control*. (2nd ed.). New York: Wiley & Son.
Wheeler, D. (2013, December 2). Gauge R&R methods compared: How do the ANOVA, AIAG, and EMP approaches differ. *Quality Digest*. Retrieved from www.qualitydigest.com
Wheeler, D., & Lyday, R. (1989). *Evaluating the measurement process*. (2nd ed.). Knoxville, TN: SPC Press.

Biographies

MUSTAFA SHRAIM is an assistant professor in the Department of Engineering Technology & Management at Ohio University in Athens, Ohio. He received both of his BS and MS degrees from Ohio University, and a PhD in Industrial Engineering from West Virginia University. He has over 20 years of industrial experience as a quality engineer, manager, and consultant in management systems and lean/Six Sigma. In addition, he coaches and mentors green and black belts on process improvement projects in both manufacturing and service. He is a certified quality engineer and a certified Six Sigma black belt by ASQ, and a certified QMS principal auditor by IRCA in London. He was elected a fellow by ASQ in 2007. Contact Dr. Shraim at shraim@ohio.edu.

WILLIAM NEIL LITTELL is is an assistant professor in the Department of Engineering Technology & Management at Ohio University in Athens, Ohio. He received his BS, MS, and PhD from Mississippi State University. He has over 15 years of experience as a consultant implementing product lifecycle management solutions for companies to improve their ability to organize data and launch new products. He is also a certified project management professional and a certified senior manufacturing specialist. Dr. Littell can be reached at littellw@ohio.edu.

ZAKI KURUPPALIL is an associate professor, assistant department chair, and Kraft Scholar in the Department of Engineering Technology and Management in Russ College of Engineering, Ohio University. His areas of expertise include lean and agile manufacturing, precision manufacturing, and production tooling. Prior to teaching, he worked as a manufacturing engineer for an aerospace supplier. Currently, he does lean coaching for Ohio University senior leadership as well as act as a project consultant for continuous improvement programs at the University. Dr. Kuruppalil can be reached at kuruppal@ohio.edu.