NON-REPLICABLE *GRR* CASE STUDY

David Benham, DaimlerChrysler Corporation

with input from the MSA workgroup:
Peter Cvetkovski, Ford Motor Company
Michael Down, General Motors Corporation
Gregory Gruska, The Third Generation, Inc.

ABSTRACT

Gage studies provide an estimate of how much of the observed process variation is due to measurement system variation. This is typically done by a methodical procedure of measuring, then re-measuring the same parts by different appraisers. This cannot be done with a non-replicable (destructive) measurement system because the measurement procedure cannot be replicated on a given part after it has been destroyed. Here a method using ANOVA is used in a case study which demonstrates one possible way to determine measurement variation in a non-replicable system.

This paper is intended to serve as additional guidelines for the analysis of measurement systems.
www.aiag.org/publications/quality/msa3.html

Non-Replicable *GRR* Case Study

BACKGROUND

Gage studies provide an estimate of how much of the observed process variation is due to measurement system variation. This is typically done by a methodical procedure of measuring, then re-measuring the same parts by different appraisers. With a standard 10-3-3[1] *GRR*, an assumption is made that between the trials and when handing off between appraisers, no physical change has occurred to the part. All appraisers in the study have the opportunity to examine the same parts. The measurements can be *replicated* between appraisers and between trials. With most situations this is a safe assumption to make.

In Figure 1, there are 10 parts to be measured, numbered from 1 to 10. As can be seen, the same part is measured by each appraiser and for each trial. A 10-2-3 format is shown for the sake of brevity and as a lead-in to the case study which follows; randomization is not shown for the sake of clarity.

	Appraiser 1			Appraiser 2		
Part #	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3
1	1	1	1	1	1	1
2	2	2	2	2	2	2
3	3	3	3	3	3	3
4	4	4	4	4	4	4
5	5	5	5	5	5	5
6	6	6	6	6	6	6
7	7	7	7	7	7	7
8	8	8	8	8	8	8
9	9	9	9	9	9	9
10	10	10	10	10	10	10

Figure 1: Example of a Standard *GRR* Layout (10-2-3)

However, there are certain instances where the measurements cannot be replicated between trials or appraisers. (The part is destroyed or somehow physically changed when it is measured – that characteristic cannot be measured again.) That measurement characteristic is said to be *non-replicable*. An example might be a destructive weld test where a weld nut is pushed off a part and the peak amount of pushout force before destruction is measured; the weld is destroyed in the process, so it cannot be measured again. So, how can a measurement systems analysis be conducted when the part is destroyed[2] during its measurement?

1 10 parts, 3 appraisers, 3 trials.

2 Not all non-replicable MSA studies necessarily involve destroying parts. In fact, the measurands need not be parts *per se*, either. For example, when tests are done on chemical processes, the chemical sample ("part") used for testing may have been altered by the test itself and the solution it was drawn from may have been from a dynamic process where the solution is in constant motion – therefore it cannot be precisely re-sampled.

Study Approach

The first thing that must be done before tackling a non-replicable *GRR* study is to ensure that all the conditions surrounding the measurement testing atmosphere are defined, standardized and controlled – appraisers should be similarly qualified and trained, lighting should be adequate and consistently controlled, work instructions should be detailed and operationally defined, environmental conditions should be controlled to an adequate degree, equipment should be properly maintained and calibrated, failure modes understood, etc. Figure 2 in the MSA manual, Measurement System Variability Cause and Effect Diagram, p. 15, and the Suggested Elements for a Measurement System Development Checklist, pp. 36 – 38, may assist in this endeavor.

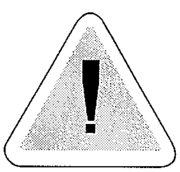
Second, there is a good deal of prerequisite work that must be done before doing a non-replicable study. The production process must be stable and the nature of its variation understood to the extent that units may be appropriately sampled for the non-replicable study – where is the process homogeneous and where is it heterogeneous? Another consideration: if the overall process appears to be stable AND CAPABLE, and all the surrounding pre-requisites have been met, it may not make sense to spend the effort to do a non-replicable study since the overall capability includes measurement error – if the total product variation and location is OK, the measurement system may be considered acceptable.

Standard *GRR* procedures and analysis methods must be changed and certain other assumptions must be made before conducting a non-replicable measurement systems analysis. The plan for sampling parts to be used in a non-replicable *GRR* needs some structure. Since the original part cannot be re-measured due to its destruction, other similar (homogeneous) parts must be chosen for the study (for the other trials and other appraisers) and an assumption must be made that they are "duplicate" or identical parts. In other words, as the "duplicate" parts are re-measured across other trials and by other appraisers, we will pretend that the same part is being measured. Refer to Figure 2. "Part 1" is now Part 1-1, 1-2, 1-3, 1-4, 1-5, 1-6, for this 10-2-3 layout. Six very similar, assumed to be identical, parts are used to represent Part 1, and so on for all 10 parts. The assumption must be made that all the parts sampled consecutively (within one batch) are identical enough that they can be treated as if they are the same. *If the particular process of interest does not satisfy this assumption, this method will not work.*

	Appraiser 1			Appraiser 2		
Part #	**Trial 1**	**Trial 2**	**Trial 3**	**Trial 1**	**Trial 2**	**Trial 3**
1A...1F	1-1	1-2	1-3	1-4	1-5	1-6
2A...2F	2-1	2-2	2-3	2-4	2-5	2-6
3A...3F	3-1	3-2	3-3	3-4	3-5	3-6
4A...4F	4-1	4-2	4-3	4-4	4-5	4-6
5A...5F	5-1	5-2	5-3	5-4	5-5	5-6
6A...6F	6-1	6-2	6-3	6-4	6-5	6-6
7A...7F	7-1	7-2	7-3	7-4	7-5	7-6
8A...8F	8-1	8-2	8-3	8-4	8-5	8-6
9A...9F	9-1	9-2	9-3	9-4	9-5	9-6
10A...10F	10-1	10-2	10-3	10-4	10-5	10-6

Figure 2: Non-Replicable *GRR* Layout[3]

Care must be taken in choosing these "duplicate" parts. Typically for the parts that represent part number 1 in a study, each "duplicate" is selected in a way that it is as much alike the original part as possible. Likewise for part number 2, and number 3, 4, 5, etc. These parts should be produced under production conditions as similar as possible. Consider the "5 M's +E"[4] and make them all as alike as possible. Generally, if parts are taken from production in a consecutive manner, this requirement is met.



However, the parts chosen to represent part number 2, for example, must be chosen to be *unlike* part number 1, part number 3, 4, 5, etc. So *between* part numbers, the 5 M's +E must be unlike each other. These differences must be forced to be between part numbers. The total number of duplicate parts selected for each row must equal the number of appraisers times the number of trials.[5] In Figure 2, groups of parts *within* each row are assumed to be identical, but groups of parts *between* rows are assumed to be different.

Part variation may be expressed as part-to-part, shift-to-shift, day-to-day, lot-to-lot, batch-to-batch, week-to-week, etc. With parts the minimum variation would be part-to-part – this represents the minimum possible amount of time between each part. When parts are not sampled consecutively (i.e., part-to-part), there is more opportunity for variation to occur – different production operators, different raw material, different components, changes in environment, etc.

So, within a row it is desirable to minimize variation by taking parts consecutively, thus representing part-to-part variation. Between rows it is desirable to maximize variation by taking parts from different lots, batches, etc. There may be economic, time or other constraints involved which will impose limits on the length of time we can wait to take

3 The necessary randomized presentation to the appraiser is not shown here for the sake of clarity.

4 Man, Machine, Material, Method, Measurement plus Environment. Measurement may seem redundant here, but there may be times where two or more "identical" measurement systems are used to gain the same information and this should be considered in any study.

5 When the source of measurement variation is thought to be due to equipment only, using different appraisers may not be required.

samples for the between-row data – the process must be run, the PPAP must be submitted, etc. When constraints arise and interfere with doing things the "right" way, the results may be subject to modified interpretation.

Another statistical assumption that must be made for this type of study is that the measurement error is normally distributed. This is a prerequisite for any *ANOVA* (Analysis of Variance).

ANOVA is a better analysis tool for a non-replicable measurement systems analysis than the average and range method. *ANOVA* has the power to examine interactions that the average range method will not catch.

> As a precautionary note, the results from this type of study will contain some process variation because the "identical" parts are not really identical. This may come into play when interpreting the results in terms of the error percentage related to process variation or tolerance. The better the methodology used to achieve an understanding of the production process and its corresponding measurement system, the more meaningful this non-replicable measurement study will be.

It is critical in a non-replicable *GRR* study that the parts be clearly identified and saved after testing. If any issues arise after the standard analysis, these parts may be needed for further exploration – e.g., microscopic examination.

A case study may serve to better illustrate this methodology.

CASE STUDY

A stamped part goes through a critical weld assembly process which must be destructively tested on an ongoing basis. The process has an in-house progressive stamping die which produces the steel stampings. This process is followed by robotic MIG welding (attaching an outside purchased steel rod to the stamping) in one of 6 different weld stations each of which has 4 parallel weld fixtures (only one of which will be used per assembly). Each weld fixture is assigned a letter designation, A through X. This process has been in production long enough that it has been studied and analyzed for stability and capability. Each of the 24 weld stations is providing a stable and capable process, however some are better than others. In an effort to improve the overall process, the measurement system would be analyzed using this non-replicable MSA methodology which had been recently introduced to the supplier.

Study Format

This study used the 10-2-3 format – 10 parts, 2 appraisers, 3 – thus requiring 60 parts total for the study. Given the complexity of the process, it was felt that the 10-2-3 would be more manageable than a 10-3-3. Although there are 24 weld fixtures, only 10 of them were used for this study and they were chosen using previously gathered data to represent the full range of the process. There was enough early confidence in the measurement system to make this judgment.

Similarity (homogeneity) *within* each row was created by taking 6 consecutively produced stampings (6 is chosen to meet the 2 appraisers x 3 trials requirement), then welding those 6 parts consecutively through the same weld fixture. Dissimilarity (heterogeneity) *between* rows was created by taking groups of 6 consecutive stampings from different coils of steel at a time separated by a few hours, then running them consecutively through a different weld fixture at a different time.

The rod component which is welded to the stamping is received in bulk and has already been determined to not play a major role in pull test variation. Therefore, in this study there was no effort made to maintain similarity and dissimilarity issues with the rod component.

Previous studies using common problem solving tools had shown that a manual positioning and clamping system used on the testing machine was appraiser dependent, so a new and better positioning system with hydraulic clamps was installed. Parts are located into the machine with positive locators and hydraulic clamps. A hook on the testing machine grabs the rod and mechanically pulls on the rod to destruction. A digital readout on the machine displays the peak pulloff force in pounds and reads to one decimal place. From this readout, the data is recorded and the failure mode noted (weld must pull metal from the stamping). Although the appraiser dependency was assumed to be resolved, this study still used two appraisers to verify that assumption.

A total of 60 parts were required to do this study. There were 10 groups of similar parts, 2 appraisers and 3 trials; 10 x 2 x 3 = 60. Parts were first gathered off the stamping operation, carefully numbered and quarantined until all 60 parts had been collected. These 10 groups of parts were selected at 3 hour intervals, over 3 days of production, in order to force some difference between each group of parts.

Then, each similar stamped group of parts was run through a different weld fixture. Parts introduced to each weld fixture were presented in random order within each group of 6.

	Appraiser 1			Appraiser 2		
Part #	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3
1-1R...1-6R	1-6R_{53}	1-5R_{25}	1-3R_{7}	1-4R_{40}	1-1R_{34}	1-2R_{32}
2-1P...2-6P	2-2P_{27}	2-6P_{35}	2-5P_{57}	2-4P_{12}	2-3P_{43}	2-1P_{17}
3-1H...3-6H	3-1H_{21}	3-2H_{36}	3-6H_{1}	3-5H_{56}	3-4H_{10}	3-3H_{26}
4-1G...4-6G	4-4G_{46}	4-6G_{42}	4-3G_{8}	4-2G_{28}	4-1G_{55}	4-5G_{30}
5-1E...5-6E	5-4E_{5}	5-3E_{20}	5-1E_{13}	5-6E_{54}	5-2E_{39}	5-5E_{50}
6-1F...6-6F	6-1F_{52}	6-3F_{3}	6-4F_{37}	6-5F_{29}	6-2F_{51}	6-6F_{45}
7-1M...7-6M	7-6M_{16}	7-4M_{11}	7-1M_{23}	7-2M_{6}	7-3M_{15}	7-5M_{14}
8-1O...8-6O	8-6O_{49}	8-3O_{60}	8-1O_{33}	8-5O_{41}	8-2O_{44}	8-4O_{19}
9-1Q...9-6Q	9-5Q_{31}	9-6Q_{59}	9-3Q_{24}	9-2Q_{4}	9-4Q_{9}	9-1Q_{2}
10-1T...10-6T	10-2T_{22}	10-5T_{18}	10-3T_{47}	10-4T_{58}	10-1T_{48}	10-6T_{38}

Figure 3: Layout Used for Case Study

Referring to Figure 3, each row shows the "similar" parts. "1-1R" stands for Stamping #1 of the first group of 6 stampings, which was run through weld fixture R. "1-2R" stands for Stamping #2 of the first group of stampings, which was run through weld fixture R. "10-1T" stands for Stamping #1 of the tenth group of stampings, which was run through weld fixture T.[6]

The parts were numbered with the identification shown in Figure 3. Parts within each row were presented in a *random order* to both the weld assembly operation and to the weld test operation, each with a different random order. The order shown above is for the weld test operation; the order for the weld assembly operation within each row was a different random order and is not shown here. Such randomization reduces the possibility of any bias present in the order of manufacturing and/or testing.

Once all assembly was completed, the parts were presented to the appraisers for destructive testing and the data were recorded. Parts were saved, preserving the original part numbers, in case any post-analysis needed to be done.

RESULTS

The data were put into a Minitab® "Gage R&R (nested)" routine which generated a nested *ANOVA*. A nested (vs. crossed) *ANOVA* is required for this type of study because all parts are not tested by (crossed with) all appraisers across all trials – they cannot be because they are destroyed after one test. Each appraiser cannot be crossed with each part. Other charts were also generated by Minitab.

The first thing to look at is the Gage Run Chart.

[6] For the sake of clarity here, the randomized order of total presentation to the appraisers running the test machine is shown as a subscript.

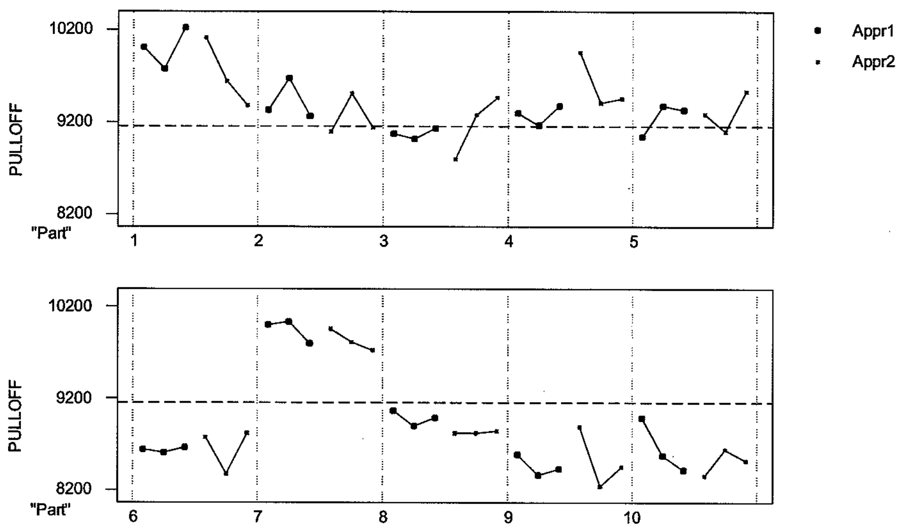


Figure 4: Gage Run Chart – Weld Tester, Rod Pulloff

The Gage Run Chart in Figure 4 shows plotted points that represent individual measurement values made on each "part" by each appraiser during each trial. The data is grouped by appraiser and presented in the order in which the trials occurred. The horizontal dashed line represents the overall mean of all the individual values displayed on the chart. The abscissa shows each "part" number (a group of 6 parts in this case).

This type of chart provides visual clues as to the presence of any patterns in the data. It is desirable that there be no particular pattern within "parts." If undesirable patterns are present, then further, more sophisticated investigation may be required. It is also desirable that there be differences shown between "parts". If these differences do not occur then the measurement system cannot distinguish between the parts used in the study. Since there are no control limits or other statistical guides displayed on this chart, some guesswork and common sense may be required when reviewing it.

In this particular study, there are no significant patterns within "parts" and there is some difference between "parts."

Next the summary of the *GRR* graphics is reviewed.

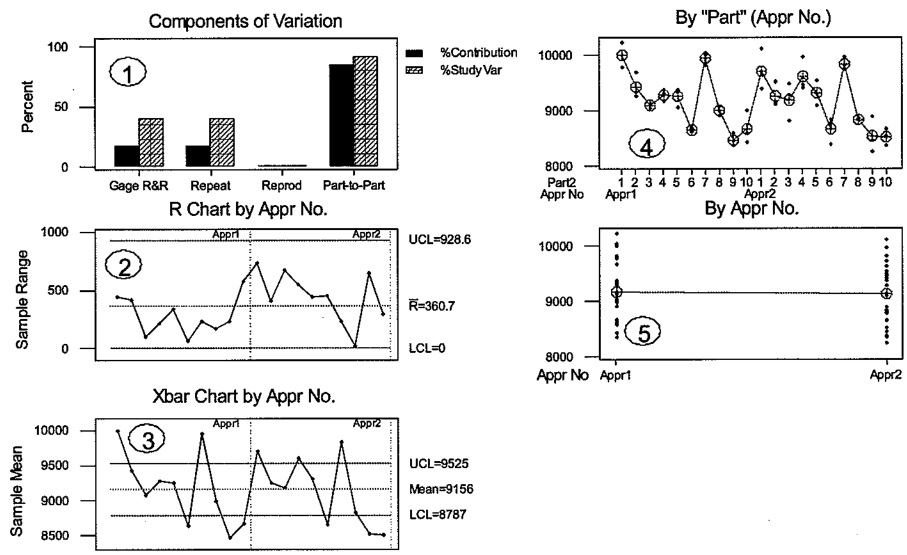


Figure 5: *GRR* Summary Graphics – *GRR* Study (Nested), Weld Tester, Rod Pulloff

A graphical summary, generated by Minitab, of this study is presented in Figure 5. Five graphs are shown and are numbered from 1 to 5. Presentation and analysis of these charts falls in line with a standard *GRR*.

Graph No. 1 shows the components of variation. Four sources are displayed here

- Gage R&R – is the variation due to the measurement system.
- Repeatability – is the variation due to the measuring equipment.
- Reproducibility – is the variation due to differences in the appraisers.
- Part-to-Part – is the variation due to measurements taken across different parts.

Two bars are shown for each source of variation[7]:

- % Contribution – is 100 times that variance component source divided by the *total variance* (σ^2).
- % Study Variation – is 100 times that study variation source divided by the total *study variation* (σ).

[7] It is possible in Minitab to also display "% Tolerance" and "% Process" but these options were not chosen here.

Graph No. 2 is a Range (*R*) Chart showing the range of the readings for each appraiser on each "part." Note that there is an upper control limit (*UCL*) and lower control limit (*LCL*) and that to be acceptable all points should be within these limits. Out of control conditions means there is some sort of inconsistency occurring and should be investigated before going further.

The data in this case study is all within the control limits and is considered acceptable with respect to the Range Chart.

Graph No. 3 is an Xbar ($\overline{X}$) Chart showing the average of each appraiser's 3 readings on each "part." Note that there is an upper control limit (*UCL*) and lower control limit (*LCL*) and that to be acceptable, approximately 50% of the points should be *outside these control limits.* The distance between the control limits represents the band of measurement system variation. If all points were within the control limits it would mean that no significant distinction can be made between any of the parts in the study.

In the data for this case study, 11 of 20 points (55%) are out of control and this is considered acceptable for the Xbar Chart.

Graph No. 4 is a chart of the part-to-part variation across the study. For this particular study, Graph 4 does not add much value.

Graph No. 5 shows the grand average for each appraiser as well as points representing the individual readings. The horizontal line between these points indicates a visual reference for the difference in the grand average – the flatter this line the less difference there is between these grand averages.

In the data for this case study, the reference line is quite flat indicating very little difference in the grand average for each appraiser.

Graphs 4 and 5 may be used as clue generators for further analysis. Also, other graphics, such as histograms, may be used for more detailed analysis.

While reviewing the graphics it is a good idea to at the same time review the ANOVA summary of the *GRR* data. Again, this summary is the same as a standard ANOVA table and what would be generated by a standard *GRR* study.

Gage R&R Study - Nested *ANOVA*

Nested *ANOVA* Table

SOURCE	DF	SS	MS	F	P
Appraiser	1	15636	15636	0.0202	0.88869
Part(Appraiser)	18	13965615	775868	16.7447	0
Repeatability	40	1853410	46335		
Total	59	15834661			

Gage R&R

SOURCE	VARCOMP	% CONTRIBUTION (OF VARCOMP)
Total Gage R&R	46335	16.00
Repeatability	46335	16.00
Reproducibility	0	0
Part-To-Part	243177	84.00
Total Variation	289513	100.00

SOURCE	STD DEVIATION (SD)	STUDY VARIATION (6 * SD)	% STUDY VARIATION (% SV)
Total Gage R&R	215.256	1291.54	40.01
Repeatability	215.256	1291.54	40.01
Reproducibility	0	0	0
Part-To-Part	493.130	2958.78	91.65
Total Variation	538.064	3228.38	100.00

CONCLUSIONS

Given all the above information one must make some decisions about the measurement system – is it acceptable, is it useable, is more study needed, does an appraiser need better training, is customer approval required for use, etc? For acceptability, generally the *GRR*% is reviewed for suitability[8].

In this particular case study example the overall *GRR*% = 40.01% which does not make for a clean interpretation. Is this acceptable given that the traditional upper limit for even a marginally acceptable measurement system is 30%?

Consider:

1. The overall process to which the *GRR*% is compared is actually represented only by the parts (and weld fixtures) chosen for this study. Only 10 of the 24 weld fixtures were included in this study. Not all of the process is necessarily represented here. If the process variation due to all the fixtures is much larger than that of the 10 selected for this study, then the measurement system may be acceptable based on MSA guidelines.
2. The data in any non-replicable study such as this will necessarily include SOME process variation. So some portion of the 40.01 *GRR*% is actually process variation. It is impossible to separate all process variation from measurement system variation with this scheme.
3. The machine used to do the destruct pulloffs is a relatively sophisticated and expensive piece of equipment. How much

[8] *ndc* is also typically reviewed at this time, however for the sake of brevity it will not be shown here.

more could it be improved and would it be cost effective? These would be questions for a possible cost study.

Acceptability Issues

Determining the acceptability of results of a non-replicable study involves many other possible issues which go beyond the immediate subject of this paper.

- Production process stability: In order to properly analyze and interpret a non-replicable measurement system study, the production process must be statistically stable. Since the measurement system stability is contained within the production process data, if the overall data demonstrates stability, then the measurement system is theoretically stable. Data collection frequency and sample size must be balanced between Average Run Length (ARL)[9] and cost of inspection.
- Production process capability: If the production process is marginally capable, there is not much breathing room between its natural process limit(s) and the measurement system's "gray area" (see discussion in Chapter III, Section C, of MSA-3). Acceptability limits (for the process data) might be used when the non-replicable measurement system analysis shows marginal or questionable acceptability. These acceptability limits would be calculated by using a 3 or 4 sigma multiplier[10] of the σ_{GRR}, then adding that value to the lower specification limit and subtracting it from the upper specification limit. However, for this to work with any degree of confidence, there must be adequate breathing room between the production process and these limits. In the example provided, where there is only a lower specification limit, the resultant value from using this multiplier would be added to the lower specification limit only; for a process with only an upper specification limit, the resultant value from using the multiplier would be subtracted from the upper specification limit only.
- Long-term measurement system stability: Some effort should be made to study the measurement system's stability on a long-term basis using techniques similar to what has been described above. These techniques may be scaled down to "mini" studies and used on an ongoing basis.
- Ironically, the better the process Cp[11], the more difficult it may be to establish acceptable non-replicable measurement system analysis results using the methods here. Recall that a major requirement for this method to work successfully is that one can knowingly produce similar (homogeneous) parts (to be used within "parts" and operator across trials) and dissimilar (heterogeneous) parts (to be used between "parts" and operators). If the production process has a very

9 The average run length (ARL) at a given quality level is the average number of samples (subgroups) taken before an action signal is given.

10 This multiplier value would be chosen depending on the amount of risk one chooses to accept.

11 Technically a process with unilateral tolerance (such as demonstrated in the case study here) has no Cp. What is meant here is a process with a relatively small amount of variation.

tight Cp, it may not be possible to meet this requirement. However, if such a process were to demonstrate stability and a high Cpk, then the measurement system is probably acceptable (provided there are no bias or linearity issues) because the data that generated this high Cpk would include measurement variation.

- Bias and linearity are not evaluated by this method. As with any *GRR* study, only repeatability and reproducibility are considered. The measurement equipment calibration plan becomes critical to the overall acceptability of the non-replicable measurement system.

Summary

As can be seen, a non-replicable measurement study may raise as many questions as it resolves. Due to the relative sophistication and expense of the measurement equipment used in this non-replicable measurement process, and due to the factors mentioned above, it was decided that this measurement process was suitable for gaging continuous improvement of the manufacturing process. At the same time, opportunities for improving the measurement system would also be evaluated in the future.

This case study is a single event and therefore has been termed a non-replicable *GRR study*. The bigger picture with any measurement *system* is the long-term, cradle-to-grave concept. The study here may be used to initially qualify a system, but more work is required to control that measurement system over time to ensure its stability and usefulness in making the appropriate process control and/or capability decisions and continuous improvement. To keep tabs on this measurement process over time, a control chart could be used to record the results of special, consecutive samples taken periodically and those results used to determine the system's stability.

Finally, the methodology above was what was chosen for this particular situation. There are other methods which will serve to determine non-replicable measurement error and each must be carefully selected to suit the particular situation. The methods shown here merely illustrate a single approach for a single situation and hopefully it will spur the reader to consider developing an approach for their own non-replicable measurement error situation.

Dave Benham works at DaimlerChrysler Corporation in Auburn Hills, Michigan, as a Senior Consultant in Supplier Development. He has a B.S. in Psychology and M.A. in Education from Michigan State University. Dave is an ASQ Certified Quality Engineer and Certified Reliability Engineer. He has worked in the automotive industry in the field of quality for 25 years.

ADDITIONAL READING

Bergeret, F.; Maubert, S.; Sourd, P.; Puel, F.; "Improving and Applying Destructive Gauge Capability," *Quality Engineering*, Vol. 14, No. 1, September 2001, pp. 59-66. (http://www.asq.org/info/library/faq/gagerr/articles.html#destructive)

Conklin, Joseph D., "Assessing Measurement Error in a Destructive Test," 47th Annual Quality Congress, May 1993, Boston MA, Vol. 0, No. 0, May 1993, pp. 400-405. (http://www.asq.org/info/library/faq/gagerr/articles.html#destructive)

Ingram, David J.; Taylor, Wayne A.; "Measurement System Analysis," Annual Quality Congress Proceedings, Philadelphia, PA, Vol. 52, No. 0, May 1998, pp. 931-941. (http://www.asq.org/info/library/faq/gagerr/articles.html#destructive)

Phillips, Aaron R.; Jeffries, Rella; Schneider, Jan; Frankoski, Stanley P.; "Using Repeatability and Reproducibility Studies to Evaluate a Destructive Test Method," *Quality Engineering*, Vol. 10, No. 2, December 1997, pp. 283-290. (http://www.asq.org/info/library/faq/gagerr/articles.html#destructive)

Spiers, Beryl, "Analysis of Destructive Measuring Systems," 43rd Annual Quality Congress, May 1989, pp. 22-27. (http://www.asq.org/info/library/faq/gagerr/articles.html#destructive)

Wheeler, D. J., "Evaluating the Measurement Process When the Testing is Destructive," TAPPI, 1990 Polymers Lamination & Coatings Conference Proceedings, pp. 805-807 (www.tappi.org; plc90905.PDF).