Journal of
INDUSTRIAL
TECHNOLOGY

Volume 23, Number 4 - October 2007 through December 2007

Gauge R&R: An Effective Methodology for Determining the Adequacy of a New Measurement System for Micron-level Metrology

by Dr. David W. Hoffa and Dr. Chad Laux

Peer-Refereed Article

KEYWORD SEARCH

Lean Manufacturing/Six Sigma
Metrology
Quality
Quality Control
Research Methods
Statistical Methods

The Official Electronic Publication of the National Association of Industrial Technology • www.nait.org

Gauge R&R: An Effective Methodology for Determining the Adequacy of a New Measurement System for Micron-level Metrology

by Dr. David W. Hoffa and Dr. Chad Laux

Dr. David Hoffa is a Quality Engineer for Accumold®, the "world leaders in Micro-Mold® manufacturing solutions," based in Ankeny, Iowa. He is responsible for continuous metrological, organizational, and product quality improvement through the application of sound, statistical methods and Lean Manufacturing concepts. He was awarded his Ph.D. in Industrial Technology from Iowa State University in 2006. He is a Certified Senior Industrial Technologist and the 2005 recipient of the NAIT Foundation's Dr. Clois E. Kicklighter Doctoral Student Scholarship Award.

Dr. Chad Laux is a Visiting Assistant Professor in the Industrial Technology Department at Purdue University in Anderson, Indiana. He teaches courses in industrial organization, statistics, and ergonomics. His research interests include advanced manufacturing, agriculture biotechnology, and quality process improvement. He is a certified Six Sigma Blackbelt from General Electric Co, Caterpillar Inc, and the American Society for Quality. He is a NAIT Certified Senior Industrial Technologist and the 2007 recipient of the Silvius-Wolansky Doctoral Fellowship Award from the Department of Agriculture and Biosystems Engineering at Iowa State University.

Abstract

As engineering tolerances approach the submicron range and customer demands for statistically verified quality levels increase, Accumold, a highly specialized, high-tech manufacturer of injection molded lead-frame and small- and micro-scale plastic parts, has realized a critical need for an improved measurement system. Many measurement technologies (including confocal and dynamic range lasers, scanning electron microscopes, white light interferometers, and precision video) from manufacturers around the world were explored. The ANOVA-based "Gauge R&R" (Repeatability and Reproducibility) study was employed as the main analytical tool for determining the efficacy of a potential measurement system; factors like potential for automation, ease of use and programming, and shop floor robustness were considered. The Gauge R&R studies effectively provided management with quantitative measures for determining whether a candidate system was adequate for measuring the critical feature of interest, which was a 10 ± 2 micron difference between adjacent step heights, and as a result, Accumold has purchased a new multi-sensor system capable of achieving a precision to tolerance ratio (a.k.a. Gauge Capability Ratio, study variance to tolerance ratio, et al.) of 11% and resolving up to 74 distinct categories when measuring these features.

Introduction

To compete in a global marketplace, manufacturers are increasingly turning to advanced manufacturing techniques to increase productivity and gain a competitive advantage. This trend requires management to be able to make decisions based on proper quantitative analysis of data. In the manufacturing process, control of variation with an increasingly high degree of precision demands an improved degree of measurement effectiveness. Measurement Systems Analysis (MSA) is a collection of statistical methods (which includes the Gauge Repeatability and Reproducibility study) for the analysis of measurement system capability (Automotive Industry Action Group [AIAG], 2002; Smith, McCrary, & Callahan, 2007).

Accumold[1], a highly specialized, high-tech manufacturer of injection molded lead-frame and small- and micro-scale plastic parts, had realized a need for an improved measurement system as a result of a new customer demand for a critical-to-function feature of a difference of 10 ± 2 microns (394 ± 79 millionths of an inch) between two adjacent step heights. This particular customer also demanded that an ANOVA-based Gauge Repeatability and Reproducibility study be performed and the result of the accompanying precision to tolerance ratio (P/T) exhibit less than 30% error, per MSA techniques

---

[1] Accumold ® is a registered trademark of Accu-Mold LLC.

(AIAG, 2002; Smith, McCrary, & Callahan, 2007).

For illustration's sake, let us first consider the requirement of 10 ± 2 µm. The thickness of a human hair is approximately 90 µm; therefore, not only is this specification asking Accumold to injection mold features into plastic parts which are 1/9th the thickness of a human hair, they are also demanding that Accumold maintain total tolerance on that feature equivalent to the thickness of one sliver of a human hair which has been split along its length 22 times!

Review of Literature

Measurement Systems Analysis

Deming once stated that knowledge of variation was one of the most powerful tools a company could apply in the quest for improvement (Joiner & Gaudard, 1990). Because variation is inherent in a process and is unpredictable, strategies to minimize variation are common in manufacturing (McGhee, 1985; MacKay & Steiner, 1997). Understanding the individual components of process variation (measurement system variation, in particular) is critical to this process because the reduction of process variation requires the ability to discriminate between process variation and measurement variation (Ishikawa, 1982; Juran, 1990; Persijn & Nuland, 1996). Measurement Systems Analysis (MSA) is based on the philosophy that measurement error masks true process capability; therefore, they are performed prior to any process improvement activities in order to quantify and minimize the measurement error (Harry & Lawson, 1992). Indeed, popular quality improvement programs such as Six Sigma utilize managing for measurement as a major analysis activity (Antony, Kumar, & Tiwari, 2005; Goffnett, 2004; Harry & Lawson, 1992; Horel, 1998; Hu, Barth, & Sears 2005; Pan, 2006). Balano's (1994) survey of quality professionals determined that containing measurement variation was the primary responsibility of quality managers in the manufacturing environment. There appears to be a gap, however, between the knowledge and practice of measurement studies and the actual deployment of measurement improvement techniques in organizations with formal quality management programs (Smith, McCrary, & Callahan, 2007).

AIAG (2002, p. 5) describes a measurement system as "a collection of instruments or gauges, standards, operations, methods, fixtures, software, personnel, environment, and assumptions used to quantify a unit of measure or the complete process used to obtain measurements". MSA quantifies 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 (AIAG, 2002, 2005). The components of measurement system variation include bias, stability, repeatability, and reproducibility, where bias is the difference between a measurement and a reference value (also known as average accuracy [accuracy is the one-time difference between a measurement result and a known standard]); stability quantifies a change in bias over time; repeatability is the variation of measurements due to instrument error (also known as precision); and reproducibility is the variability resulting from external sources such as operators and their unique techniques, setups, and environmental fluctuations over time (AIAG 2002; Engel & De Vries, 1997; Smith, McCrary, & Callahan, 2007). A Gauge Repeatability & Reproducibility (GR&R) study estimates the repeatability and reproducibility components of measurement system variation with the primary objective of assessing whether the gauge is appropriate for the intended application (AIAG, 2002; Burdick, Park, & Montgomery, 2005). Assessing the suitability of the examined measurement systems using the GR&R study was the primary goal of this paper.

Gauge Repeatability & Reproducibility

GR&R is a well-covered topic in the literature (AIAG, 2002; Burdick, Borror, & Montgomery, 2003; Dolezal, Burdick, & Birch, 1998; Goffnet, 2004; Montgomery & Runger, 1993a, 1993b; Pan, 2004, 2006; Persijn & Nuland, 1996; Smith, McCrary, & Callahan, 2007; Vardeman & Job, 1999). Though a detailed description of GR&R is beyond the scope of this paper, a brief review of the application is appropriate.

Repeatability ($\hat{\sigma}_{repeatability}$) can be determined by measuring a part several times, effectively quantifying the variability in a measurement system resulting from the gauge itself (AIAG, 2002; Smith, McCrary, & Callahan, 2007; Pan, 2006). This can also be thought of as "within operator" variability (Smith, McCrary, & Callahan, 2007).

Reproducibility ($\hat{\sigma}_{reproducibility}$) is determined from the variability created by several operators measuring a part several times each, effectively quantifying the variation in a measurement system resulting from the operators of the gauge and environmental factors such as time (AIAG, 2002; Burdick et. al, 2003; Pan, 2006; Tsai, 1989). This can also be thought of as "between operator" variation (Smith, McCrary, & Callahan, 2007; Pan, 2004).

"Total Gauge R&R" ($\hat{\sigma}_{R\&R}$) is the estimate of the combined estimated variation from repeatability and reproducibility (AIAG, 2002).

Total measurement system variation is the sum of the variation of Total Gauge R&R ($\hat{\sigma}^2_{R\&R}$) with part-to-part variation ($\hat{\sigma}^2_{part}$: the variability of the individual pieces) (AIAG, 2002; Pan 2006).

These individual sources of variation combine as shown in the two-way random effects model in Equation 1, where y_{ijk} = the kth measurement on part i by operator j; where μ the mean of all measurements taken by all operators (an unknown constant); where α_i is the random effect of the different parts (normal random variables with mean 0 and variance $\hat{\sigma}^2_{\alpha}$); where β_j is the random effect of the different operators (normal random variables with mean 0 and variance = $\hat{\sigma}^2_{\beta}$); where $\alpha\beta_{ij}$ is the random interaction effect between certain part and operator combinations (a normal random variable with mean 0 and variance $\alpha\beta^2_{ij}$); and where ε_{ijk} is measurement error (a normal random variable with mean 0 and variance σ^2). These variance components are the individual sources of variation in a measurement system (Vardeman & Job, 1999, p. 21-22).

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk} \qquad (1)$$

There are two primary methods for conducting a GR&R study, each based upon how measurement variation is calculated. Mandel (1972) first described quantifying measurement variation by utilizing ANOVA. This is prescribed by Pan (2004, 2006) since the variability component created by the interaction of inspectors and parts can be determined. The average and range method described by Burdick et. al. (2003), Hart (2005), Montgomery and Runger (1993a), Pan, (2006), and Vardeman and Job (1999) uses a range chart to ensure stability of the measurement process and is easier to compute without the aid of statistical software, but it can not reveal operator-by-part interactions (ignoring a significant interaction effect could provide an overly-optimistic result) (AIAG, 2002; Pan, 2004).

Statement of the Problem

As a result of customer demands for ever-smaller injection molded plastic part features with increasingly tight tolerances, Accumold's pre-study metrology capabilities had been exceeded, creating the need for a more capable measurement system.

Purposes

This study served two purposes: first, to identify a new measurement system capable of providing Accumold with three axes of repeatable, reliable, automatable measurements of precision, injection molded plastic parts with engineering tolerances as tight as ± 2 µm; and second, to illustrate the practicality of the GR&R study as a decision-making tool for Industrial Technology students and practitioners.

Methodology

Four measurement systems from three vendors were deemed potential candidates for satisfying Accumold's requirements and were tested. In order to avoid any issues with publicizing the systems' names, they shall be referred to only as "WLI I", "WLI II", and "Multi-sensor".

The main statistical analysis tool utilized in this study to determine the suitability of a measurement system for measuring micro-scale features on injection molded plastic parts with engineering tolerances as tight as ± 2 µm was the ANOVA-based GR&R study as provided by the *Gage R&R study (crossed)* module in Minitab release 13.32 (2000). The factor of interest was the Z-axis distance between the two adjacent step height features, and the response was that Z-axis difference in height in microns (or fractions of millimeters).

Ten parts, three operators, and three trials are typical for both ANOVA and average and range methods (AIAG 2002); however, Burdick et al. (2005) state that traditional designs are not sufficient to discriminate between good and bad parts and recommend a minimum of six operators. On the contrary, Pan (2004) states that the total number of measurements should be determined first based upon cost and subsequently, by determining the combination of operator and replicates. In the studies discussed here, setup time for the measurement systems was a constraint that mandated a smaller number of measurement runs involving fewer operators; the first three GR&R studies discussed here included only two operators measuring each of 10 parts twice, which resulted in a data set comprised of 40 individual measurement results (see Table 1), with the final study using the typical 3/10/10 design resulting in 90 measurements.

Precision to tolerance ratio or P/T (also known as Gauge Capability Ratio, study variance to tolerance ratio, et al.) is computed according to Equation 2 (Minitab performs this computation and places the result in the "Total Gage R&R" row under the "% Tolerance" column) and can be used to determine the suitability of a measurement system for the application (Minitab, 2000; Pan, 2006).

$$P/T = \frac{k \times \hat{\sigma}_{R\&R}}{USL - LSL} \qquad (2)$$

Burdick et al. recommend setting *k* to 5.15 or 6 (5.15 was used for all of the following analyses):

> *The value k = 6 corresponds to the number of standard deviations between "natural" tolerance limits of a normal process. The value k = 5.15 corresponds to the limiting value of the number of standard deviations between bounds of a 95% tolerance interval that contains at least 99 % of a normal population (2003).*

According to AIAG (2002), if the P/T is under 10%, the measurement system is acceptable; if between 10% and 30%, a measurement system "may be acceptable based upon the importance of application" (p. 77) and the associated expenses; and if over 30%, the

Table 1. Parameters of Accomold Gauge R&R Studies

System	Parts (i)	Operators (j)	Replicates (k)	Total (n)
WLI I	10	2	2	40
WLI II	10	2	2	40
Multi-sensor A	10	2	2	40
Multi-sensor B	10	3	3	90

measurement system is considered unacceptable.

Another statistic provided by Minitab (2000) as a part of a GR&R study and recommended by AIAG (2002) is the "number of distinct categories". Also known as classification ratio, this statistic is computed by dividing $\hat{\sigma}_{part}$ by $\hat{\sigma}_{R\&R}$, multiplying by 1.41, and rounding down to the nearest whole number (AIAG, 2002; Wheeler & Lyday, 1989). AIAG (2002) states that if the measurement system resolves less than two distinct categories, the data is all noise – the system is of no value; two distinct categories divide the data into high and low groups (essentially reducing the variable data to attribute data); three distinct categories indicates that a measurement system is capable of defining low, middle, and high groups; and four distinct categories "would be much better" than 0-3 (p. 163). The measurement system is adequate if the number of distinct categories is greater than or equal to five (AIAG, 2002, p. 117).

Based on expected customer demands, Accumold was requiring an ANOVA-based GR&R study with a resulting P/T less than 30% and a minimum of two distinct categories (see *Multi-sensor* section) in order to consider a measurement system suitable for purchase.

A gauge block with a 90° corner was temporarily affixed to the stage in each test and operators were instructed to butt the part into the corner. This is obviously less than optimal, but was all that was available for testing and an improvement over no fixture whatsoever.

Sampling

The number of parts in these studies is based upon common practice and the practical constraints of the process. Two samples were gathered for the four individual GR&R studies; the first sample was comprised of 10 pre-production parts gathered at one time from a single cavity of a four cavity mold, and the second was comprised of 10 pre-production parts gathered from each of the four mold cavities across three months of "retains" to capture a potentially broader range of the process variability (AIAG, 2002; Burdick et al., 2003; Pan, 2004, 2006). This second sample is discussed in greater detail in the *Multi-sensor B* section of the *Findings*. Due to the proprietary nature of these parts, no specifics beyond what is mentioned here can be given.

WLI I

WLI I was a white light interferometer with published submicron-level accuracy and repeatability. This system had a motorized stage and a selection of "pens" which provided a broad range of working distances and accuracy/repeatability. As previously stated, the unavoidable need to compress the duration of this study required Accumold to gather the data using two operators, 10 parts, and two trials. It was discovered during testing that WLI I was critically limited in that it was not designed specifically for measuring dimensions in the X-Y plane – the intended purpose was profilometry.

WLI II

This system was similar to WLI I in that they shared the same base technology. The major difference between WLI I and WLI II was that WLI II's software was better suited for three-axis dimensional metrology.

Multi-sensor

Two separate GR&R studies were conducted on two different multi-sensor base platforms (*Multi-sensor A* and *B*), both using the identical laser probe (the design of the probe allows for its removal and installation on any machine equipped for its use). The two base systems use the same software and differ in terms of appearance and degree of precision (Multi-sensor B is more accurate and repeatable than Multi-sensor A), but they were functionally identical. This is discussed in full detail in the *Findings* section.

The Multi-sensor systems were essentially live video microscopes with three powered axes, zoom lens, and available laser, white light confocal, and contact measurement probes. Among these, the most precise was the white light confocal probe; unfortunately, the surface of the part being tested was too diffuse for this probe to detect. As a result, the laser probe was used for the critical Z-axis measurements.

X-Y plane measurements were gathered with the camera system. This was achieved by magnifying the area of interest (in this case, a through-hole on either end of the part) to fill the viewing area of the screen and using the edge- and feature-finding tools in the metrology software to take measurements. The factor of interest here was the width between the centers of the through-holes, which was studied during the Multi-sensor B test, and the response was width in millimeters.

The Multi-sensor systems have two distinct advantages over the WLI systems: 1) the intended use is dimensional metrology (this seemed to be an afterthought on the WLI systems), and therefore, if the operator can see the feature, it can be measured; and 2) the software is capable of locating features within a window of positioning error and correcting for misalignment, making perfect fixturing unnecessary.

Findings

WLI I

The P/T for this system was 68.64%, more than doubling the 30% maximum suitable result. This came as a bit of a surprise, as the published figures for this system indicated that it should have been capable of passing the test with ease, using the rule-of-thumb 10:1 ratio (minimum) of engineering tolerance to gauge precision. The study detected less than 1 distinct category, which again is to be interpreted as the system being of no value for measuring these parts with the ± 2 μm tolerances. See Table 2 (on page 6).

WLI II

Data gathering for this system was halted as it became apparent that measurement variation was excessive; the GR&R study was not performed. Again, this variability came as a surprise because the published accuracy and repeatability of the system far surpassed the 10:1 rule-of-thumb. Additional analyses of the part with this system revealed that the surface irregularity of the parts' Electrical Discharge Machining finish (specified by the customer) was causing the measurements to vary, thereby causing poor accuracy and repeatability. Either additional software features not available at the time of testing or perfect fixturing may have alleviated this condition. In fact, taping a single part to the stage (perfect fixturing) and measuring the feature 10 times decreased $\hat{\sigma}_{repeatability}$ to less than 1 nm, although accuracy was still a problem (see Table 3). Since perfect fixturing is unrealistic, the additional software features would need to be tested for the ability to automatically correct for stage misalignment.

Table 2. Gauge R&R for WLI I

Source	StdDev (SD)	Study Var (5.15*SD)	% Study Var (%SV)	P/T (SV/Toler)
Total Gage R&R	5.33E-04	2.75E-03	97.89	**68.64**
Repeatability	4.12E-04	2.12E-03	75.59	53
Reproducibility	3.39E-04	1.74E-03	62.2	43.61
Operator	2.31E-04	1.19E-03	42.47	29.78
Operator*Part	2.47E-04	1.27E-03	45.44	31.86
Part-To-Part	1.11E-04	5.73E-04	20.43	14.32
Total Variation	5.45E-04	2.80E-03	100	70.12

Number of Distinct Categories = 0

Multi-sensor

Two tests were conducted on two different Multi-sensor base platforms, both utilizing the same optional laser probe – these will be called "Multi-sensor A" and "Multi-sensor B".

Multi-sensor A

Refer to Table 4 (see page 7). The P/T for this study was 14.65%, less than half of the 30% maximum; however, the system was only able to resolve one distinct category, which is to be interpreted as the system having no value for measuring these parts – but was this the really the case? The sample used for this test was identical to the sample used in the previous tests (10 parts pulled from a single mold cavity at one time) and since Accumold was confident in the repeatability of their processes (for example, the lead author measured a 20-piece, random sample [of a different product] taken from a single lot with the customer-supplied gauge pin measurement system and found zero variation in the feature at the micron scale), it was possible that these parts were actually too similar for the system to differentiate between them.

Multi-sensor B

To clear up the ambiguity of the first Multi-sensor test, Accumold gathered a second, more diverse sample and conducted a second GR&R study with the typical study design (10 parts, three operators, and three trials – 90 measurements in total). This sample contained parts pulled from each of the four cavities and from three distinct pre-production runs. After much discussion, it was also decided that the sample would include two parts from a prior revision of the mold where the nominal difference between step heights was 14 μm; the justification for this was that their inclusion in the sample would provide upfront certainty that there were indeed two distinct categories of parts. If the system could not make a distinction between the parts, a result of one distinct category could be accepted as accurate and the system could be deemed unsuitable.

Since it was also determined (based on sales literature) that the Multi-sensor A system was a step down from the current video measurement system capabilities, the manufacturer's highest-performing base system (again, with the same laser probe) was used. Concerns that changing base systems introduced a confounding factor into the study can be ignored because the purposes of the second test were to examine a more diverse sample to discern more distinct categories and to examine the system most likely to be purchased – the purpose was not to determine whether the systems were statistically different.

Table 3. Results of Fixed Part Repeatability Test, WLI II

Msmt #	Result (μm)
1	-9.469
2	-9.469
3	-9.469
4	-9.469
5	-9.469
6	-9.469
7	-9.469
8	-9.469
9	-9.469
10	-9.469
Range	0
Mean	-9.469
$\hat{\sigma}_{repeatability}$	0

The results of the second test of the step height feature (see Table 5, on page 7) revealed a P/T of 10.47% (approximately 4% lower [a 29% improvement] than Multi-sensor A). The majority of this difference was most likely attributable to the larger data set since both systems used the identical laser probe. Seventy-four distinct categories were resolved, indicating a system capable of differentiating these parts, and was the result of having $\hat{\sigma}_{part}$ two orders of magnitude larger than $\hat{\sigma}_{R\&R}$. Keeping in mind that AIAG (2002) states that a measurement system must resolve only five distinct categories to be considered acceptable, these results indicate that this system was suitable for use in measuring this critical feature.

Table 4. Gauge R&R for Multi-sensor A, Z-height

Source	StdDev (SD)	Study Var (5.15*SD)	% Study Var (%SV)	P/T (SV/Toler)
Total Gage R&R	1.14E-04	5.86E-04	88.75	**14.65**
Repeatability	7.44E-05	3.83E-04	58	9.57
Reproducibility	8.61E-05	4.44E-04	67.18	11.09
Operator[a]	0.00E+00	0.00E+00	0	0
Operator*Part	8.61E-05	4.44E-04	67.18	11.09
Part-To-Part	5.91E-05	3.04E-04	46.08	7.61
Total Variation	1.28E-04	6.60E-04	100	16.51

Number of Distinct Categories = 1

[a] The metrology software completely eliminated variation due to the three operators.

Table 5. Gauge R&R for Multi-sensor B, Z-height (full sample)

Source	StdDev (SD)	Study Var (5.15*SD)	% Study Var (%SV)	P/T (SV/Toler)
Total Gage R&R	8.13E-05	4.19E-04	1.91	**10.47**
Repeatability	5.95E-05	3.06E-04	1.39	7.66
Reproducibility	5.55E-05	2.86E-04	1.3	7.14
Operator[a]	0.00E+00	0.00E+00	0	0
Operator*Part	5.55E-05	2.86E-04	1.3	7.14
Part-To-Part	4.27E-03	2.20E-02	99.98	549.51
Total Variation	4.27E-03	2.20E-02	100	549.61

Number of Distinct Categories = 74

[a] The metrology software completely eliminated variation due to the three operators.

Since there was some doubt concerning the validity of including two intentionally different parts in the sample, they were removed from the model and the analysis was repeated on the remaining eight parts from the same mold revision (see Table 6 on page 8). The P/T increased slightly to 11.40% and was likely attributable to the smaller sample. Twelve distinct categories were resolved – while this was certainly a much smaller value than when using the full sample, it was still more than double the AIAG-specified value of five.

For the width between the centers of the through-holes on either end of the part (again, measured using the camera system – not the laser probe) for all 10 parts, the P/T was revealed to be 24.89% (marginally below the 30% maximum) and 16 distinct categories were resolved (see Table 7 on page 8). With the two "old revision" parts discarded, the P/T increased slightly to 26.31% and the number of distinct categories decreased slightly to 12 (refer to Table 8). Although these results approach the cutoff mark at 30%, Accumold believes that these values will improve with additional testing and options. It was observed during testing that the measurement routine was not optimized, as the system picked an incorrect edge on several occasions – this can likely be corrected in the future. Furthermore, the test system was not equipped with the optional high-resolution, monochromatic optics, which Accumold has purchased.

Implications

The outcomes of this study have direct implications for Accumold and more general implications for the readers of the Journal of Industrial Technology. The immediate implications of this research are the improved measurement capabilities for Accumold – the results of this research have successfully identified a new measurement system that will allow Accumold to ensure measurement validity on injection molded plastic part features with engineering tolerances as tight as ± 2 µm. Additionally, the new system also provides Accumold with profiling and contouring capabilities, functions that were previously unavailable in-house. It is important to note that despite the indication in the sales literature for both WLI systems that they were suitable for the task, neither performed to expectations when subjected to the GR&R test – conducting the GR&R tests on these systems effectively saved Accumold from making a $150,000 mistake.

In broader terms, this research demonstrates the practicality and effectiveness of the GR&R study as a decision-making tool; in this instance, to aid management in making a sound decision in the purchase of a new measurement system with a prescribed degree of precision. This statistical tool can be applied to any measurement system from the most advanced multi-sensor system to the simplest measuring stick. It can be used to establish a baseline of precision of a given process's measurement system, or to quantify a suspected inadequacy in a measurement system so that the process of identifying a new measurement procedure or system can begin.

Recommendations and Discussion

A company wishing to conduct GR&R analysis on a gauge will need to weigh the pros and cons of the two major types, ANOVA-based and average/range-based. ANOVA-based GR&R studies have the ability to detect any operator*part interaction that may exist, but at the cost of more difficult computations and perhaps the expense of a statistical software package. Average/range-based GR&R studies can be computed in an Excel spreadsheet (many templates are available for free on the Internet), but at the expense of never detecting a potential operator*part interaction and providing an overly-optimistic result where one exists. For illustration, Table 9 shows that the outcome of the study would have been noticeably more favorable in terms of P/T and number of distinct categories when using the average/range method (compare to Table 8).

Using as large a data set as time and money will allow is recommended. The impact of the smaller data set created by removing the two dissimilar parts is apparent when examining the results of the two Multi-sensor B tests – in each case, P/T increases and the number of distinct categories decreases (compare Table 4 to 5, 6 to 7).

The practitioners' knowledge of the appropriate MSA techniques allowed Accumold to learn about the appropriateness of several highly technical measurement systems for their unique needs in a novel, scientific way. Callahan, Amos, and Strong (2004) indicated the importance of industrial technology professionals possessing knowledge and skills in metrology and MSA techniques, and this paper corroborates that assertion. As the complexity of manufacturing processes continues to grow, individuals who can solve problems and make decisions based on quantitative methods such as MSA will become increasingly valuable to such organizations as the metrological demands of their customers change (Smith, McCrary, & Callahan, 2007).

Table 6. Gauge R&R for Multi-sensor B, Z-height (adjusted sample)

Source	StdDev (SD)	Study Var (5.15*SD)	% Study Var (%SV)	P/T (SV/Toler)
Total Gage R&R	8.86E-05	4.56E-04	12.01	**11.4**
Repeatability	6.34E-05	3.27E-04	8.6	8.17
Reproducibility	6.18E-05	3.18E-04	8.38	7.96
Operator[a]	0.00E+00	0.00E+00	0	0
Operator*Part	6.18E-05	3.18E-04	8.38	7.96
Part-To-Part	7.32E-04	3.77E-03	99.28	94.25
Total Variation	7.37E-04	3.80E-03	100	94.94

Number of Distinct Categories = 12

[a] The metrology software completely eliminated variation due to the three operators.

Table 7. Gauge R&R for Multi-sensor B, Width (full sample)

Source	StdDev (SD)	Study Var (5.15*SD)	% Study Var (%SV)	P/T (SV/Toler)
Total Gage R&R	1.93E-04	9.96E-04	8.93	**24.89**
Repeatability	1.52E-04	7.82E-04	7.02	19.55
Reproducibility	1.20E-04	6.16E-04	5.53	15.41
Operator	6.49E-05	3.34E-04	3	8.35
Operator*Part	1.01E-04	5.18E-04	4.64	12.94
Part-To-Part	2.16E-03	1.11E-02	99.6	277.59
Total Variation	2.16E-03	1.11E-02	100	278.71

Number of Distinct Categories = 16

Table 8. Gauge R&R for Multi-sensor B, Width (adjusted sample)

Source	StdDev (SD)	Study Var (5.15*SD)	% Study Var (%SV)	P/T (SV/Toler)
Total Gage R&R	2.04E-04	1.05E-03	11.38	**26.31**
Repeatability	1.60E-04	8.25E-04	8.93	20.63
Reproducibility	1.27E-04	6.53E-04	7.06	16.33
Operator	8.60E-05	4.43E-04	4.79	11.08
Operator*Part	9.32E-05	4.80E-04	5.19	12
Part-To-Part	1.78E-03	9.19E-03	99.35	229.66
Total Variation	1.80E-03	9.25E-03	100	231.16

Number of Distinct Categories = 12

Table 9. Average/range-based Gauge R&R for Multi-sensor B, Width (Adjusted sample)

Source	StdDev (SD)	Study Var (5.15*SD)	% Study Var (%SV)	P/T (SV/Toler)
Total Gage R&R	1.70E-04	8.75E-04	8.35	**21.88**
Repeatability	1.39E-04	7.15E-04	6.82	17.87
Reproducibility	9.81E-05	5.05E-04	4.82	12.63
Part-to-Part	2.03E-03	1.04E-02	99.65	261.03
Total Variation	2.03E-03	1.05E-02	100	261.94

Number of Distinct Categories = 17

References

Antony, J., Kumar, M., & Tiwari, M. K. (2005). An application of six sigma methodology to reduce the engine-overheating problem in an automotive company. *Journal of Engineering Manufacture, B8,* 14, 633-646.

Automotive Industry Action Group (AIAG). (2002). *Measurement systems analysis* (3rd ed.). Southfield, MI: Author.

Automotive Industry Action Group (AIAG). (2005). *Statistical Process Control* (2nd ed.). Southfield, MI: Author.

Automotive Industry Action Group (AIAG). (2006). *Production Part Approval Process* (4th ed.). Southfield, MI: Author.

Balano, R. M. (1994). The ten commandments of quality. *Quality Progress, 27*, 1, 41-42.

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

Burdick, R. K., Park, Y. J., & Montgomery, D. C. (2005). Confidence intervals for misclassification rates in a gauge R&R study. *Journal of Quality Technology, 37*, 4, 294-303.

Callahan, R. N., Amos, S. J., & Strong, S. D. (2004). Enhancing entry level skills in quality assurance: An industry perspective. *Journal of Industrial Technology, 20,* 4, 1-6.

Dolezal, K. K., Burdick, R. K. & Birch, N. J. (1998). Analysis of a two-factor R&R study with fixed operators. *Journal of Quality Technology, 30*, 2, 163-170.

Engel, J. & De Vries, B. (1997). Evaluating a well-known criterion for measurement precision. *Journal of Quality Technology, 29*, 4, 469-476.

Goffnett, S. P. (2004). Understanding six sigma: Implications for industry and education. *Journal of Industrial Technology, 20*, 4, 2-10.

Harry, M. J., & Lawson, J. R. (1992). *Six sigma producibility analysis and process characterization.* New York: Addison-Wesley.

Horel, R. W. (1998). Six sigma and the future of the quality profession. *Quality Progress, 31*, 6, 35-41.

Hart, M. (2005). Learning by doing: A series of hands-on projects for SPC. *Quality Engineering, 17,* 1, 172-137.

Hu, M., Barth, B., & Sears, R. (2005). Leveraging six sigma disciplines to drive improvement. *International Journal of Six Sigma and Competitive Advantage, 1*, 2, 121-133.

Ishikawa, K. (1982). *Guide to quality control.* White Plains, NY: Quality Resources.

Joiner, B. L., & Gaudard, M. A. (1990). Variation, management, and W. Edwards Deming. *Quality Progress, 23*, 12, 29-36.

Juran, J. M. (1990). *Quality control handbook* (4th ed.). New York: McGraw Hill.

MacKay, R. J., & Steiner, S. H. (1997). Strategies for variability reduction. *Quality Engineering, 10*, 1, 125-136.

Mandel, J. (1972). Repeatability and reproducibility. *Journal of Quality Technology, 4*, 2, 74-85.

McGhee, J. W. (1985). *Introductory statistics.* St. Paul, MN: West.

Minitab Inc. (2000). Minitab (Version 13.32) [Computer software]. State College, PA: Minitab Inc.

Montgomery, D. C., & Runger, G. C. (1993a). Gauge capability analysis and designed experiments. Part I: Basic methods, *Quality Engineering,* 6, 1, 115-135.

Montgomery, D. C., & Runger, G. C. (1993b). Gauge capability analysis and designed experiments. Part II: Experimental design models and variance component estimation, *Quality Engineering,* 6, 2, 289-305.

Pan, J. H. (2004). Determination of the optimal allocation of parameters for gauge repeatability and reproducibility study. *International Journal of Quality and Reliability Management, 21*, 6, 672-682.

Pan, J. H. (2006). Evaluating the gauge repeatability and reproducibility for different industries. *Quality and Quantity, 40,* 4, 499-518.

Persijn, M., & Nuland, Y. V. (1996). Relation between measurement system capability and process capability. *Quality Engineering, 9*, 1, 95-97.

Smith, R. R., McCrary, S. W., & Callahan, R. N. (2007). Gauge repeatability and reproducibility studies and measurement system analysis: A multimethod exploration of the state of practice. *Journal of Quality Technology, 23*, 1, 1-11.

Tsai, P. (1989). Variable gauge repeatability and reproducibility study using the analysis of variance method. *Quality Engineering, 1*, 1, 107-115.

Vardeman, S. B., & Job, J. M. (1999). *Statistical quality assurance methods for engineers*. New York: John Wiley & Sons, Inc.

Wheeler, D. J., & Lyday, R. W. (1989). *Evaluating the measurement process.* Knoxville, TN: SPC Press.