

Measurement System Analysis Beyond the Gage R&R

The crossed Gage R&R is where most MSA programs begin and end. Bias, linearity, stability, resolution, nested studies, and attribute agreement are where the real problems hide.

Ref: AIAG MSA Manual, 4th Ed. · Wheeler EMP III · ISO 22514-7 · ASTM E2782

I — Publication

Measurement System Analysis Beyond the Gage R&R

The crossed Gage R&R is where most MSA programs begin and end. Bias, linearity, stability, resolution, nested studies, and attribute agreement are where the real problems hide.

Ref: AIAG MSA Manual, 4th Ed. · Wheeler EMP III · ISO 22514-7 · ASTM E2782

Date 2026-04-01

Author Eric Wolters

Reading 14 min

MSA · Measurement

II — The Crossed Gage R&R

Standard Design

The standard Gage R&R uses a crossed design: each operator measures each part multiple times. A typical study uses 3 operators, 10 parts, and 2-3 trials per operator-part combination, for a total of 60-90 measurements. The design is crossed because every operator sees every part. This allows the ANOVA to decompose total variation into repeatability (within-operator, trial-to-trial), reproducibility (between-operator), the operator-by-part interaction, and part-to-part variation.

ANOVA Method vs. Range Method

The range method (also called the X-bar/R method) estimates repeatability from the average range within each operator-part cell and reproducibility from the range of operator averages. It is simpler to compute by hand but does not separate the operator-by-part interaction from reproducibility. The ANOVA method decomposes the variation completely, isolates the interaction term, and provides F-tests for each component. Use the ANOVA method. The range method is a computational shortcut from the pre-computer era. It discards information.

Interpreting the Results

%GRR (Study Variation): The measurement system variation as a percentage of the total observed variation (or the tolerance, depending on the basis). Below 10%: acceptable. 10-30%: may be acceptable depending on the application. Above 30%: unacceptable. These thresholds are from the AIAG MSA manual and are widely used, though they are not universal standards.

%Contribution: The percentage of total variance attributable to GRR. This is the squared version of %Study Variation. A %Study Variation of 30% corresponds to a %Contribution of 9%—seemingly better, but describing the same measurement system.

Number of Distinct Categories (ndc): The number of non-overlapping confidence intervals that span the part-to-part variation. Calculated as $ndc = 1.41 \times (PV / GRR)$, where PV is the part variation and GRR is the measurement system variation. $ndc \geq 5$ is the AIAG threshold for an adequate measurement system. Below 5, the measurement system cannot reliably distinguish between parts, and process control charts based on these measurements will show excessive noise.

III — Beyond the Crossed Design

Nested Studies for Destructive Testing

When measurement destroys the part (tensile testing, chemical analysis, weld peel tests), the same part cannot be measured by multiple operators. The crossed design is impossible. A nested design assigns different parts to each operator, with parts drawn from the same homogeneous batch. The between-operator variation is confounded with the between-part variation within batches, so the study relies on the assumption that parts within a batch are interchangeable. Select parts from a well-mixed, homogeneous source. Label them carefully. If the batch is not truly homogeneous, the study attributes material variation to operator variation.

Attribute Agreement Analysis

When the measurement is a classification (pass/fail, grade A/B/C, defect type), the standard Gage R&R does not apply. Attribute agreement analysis assesses whether appraisers agree with each other and with a known standard. **Cohen's kappa** measures pairwise agreement corrected for chance. **Kendall's coefficient of concordance** measures agreement among multiple raters on ordinal scales. **Fleiss' kappa** extends to multiple raters with nominal categories.

Attribute measurement systems are typically less capable than variable systems. A visual inspection station that classifies surface defects will have lower agreement than a CMM measuring a bore diameter. When attribute agreement is poor, the first question is whether the operational definitions are clear. "Acceptable surface finish" means different things to different inspectors unless the standard includes reference samples, photographs, and unambiguous boundary conditions.

IV — Bias and Linearity

Systematic Error vs. Random Error

The Gage R&R measures precision (random error). Bias measures accuracy (systematic error). A measurement system can have excellent precision (%GRR = 5%) and terrible accuracy (reads 0.02 mm high across the entire range). The Gage R&R will not detect this. Bias is assessed by measuring a reference standard of known value multiple times and comparing the average measurement to the reference. If the difference is statistically significant, the gage is biased.

Linearity Across the Measurement Range

Linearity studies assess whether bias changes across the measurement range. A gage that reads 0.01 mm high at the low end and 0.03 mm high at the high end has non-zero linearity. This is common in mechanical gages (spring-loaded indicators), optical systems, and any sensor with nonlinear response characteristics.

The linearity study measures reference standards at 5+ points spanning the operating range. The bias at each point is plotted against the reference value. If the relationship is flat, the bias is constant and can be corrected by a simple offset. If the relationship has a slope, the gage needs calibration adjustment or replacement.

How Bias Invalidates Capability Differently

Random measurement error (repeatability + reproducibility) inflates the observed process variation and makes capability look worse than it is. Bias shifts the apparent process mean without changing the apparent spread. A biased measurement system can produce a Cpk that looks centered and capable when the actual process is off-center. Conversely, it can produce a Cpk that looks off-center when the process is fine. Bias distorts the centering assessment; random error distorts the spread assessment. Both matter; they distort differently.

V — Stability and Resolution

Measurement Drift

A gage that was accurate at calibration may drift over time. Temperature changes, mechanical wear, battery degradation, contamination of optical surfaces, and reference standard drift all contribute. A stability study plots the measurement of a control standard over time on a control chart. If the chart shows trends or shifts, the calibration interval is too long or the gage is degrading. Stability studies are inexpensive: measure one reference standard at the start of each shift and plot the results. The information value is high relative to the cost.

Resolution Requirements

Resolution is the smallest increment the measurement system can detect. The AIAG guideline is that the measurement resolution should be at most one-tenth of the specification tolerance or one-tenth

of the total process variation, whichever is smaller. This is the "10-bucket rule": the measurement system should divide the tolerance into at least 10 distinguishable levels.

The 10-bucket rule is sometimes too conservative. For attribute-type measurements or for processes where the tolerance is very tight relative to the achievable resolution, 5 buckets may be acceptable. For SPC applications where the chart needs to detect small shifts, 10 buckets is a minimum. The key question is whether the measurement resolution limits your ability to detect the changes you need to detect. If the process varies by 0.003 mm and the gage resolution is 0.001 mm, you have 3 buckets—the control chart will look like a staircase and run rules become unreliable.

VI — When %GRR Lies

Inflated %GRR from Poor Part Selection

%GRR (Study Variation) is a ratio: measurement variation divided by total variation. If the parts selected for the study span a narrow range (all near nominal), the total variation is small and %GRR is inflated. The same measurement system evaluated with parts spanning the full tolerance range would show a smaller %GRR. This is not a gage problem; it is a study design problem. Select parts that represent the full range of process variation. The AIAG manual recommends choosing parts that span 80-100% of the expected process range.

The Interaction Term

A significant operator-by-part interaction means that certain operators measure certain parts differently than expected from the main effects alone. This can indicate that some parts are harder to measure consistently (complex geometry, inconsistent fixturing) or that operators use different techniques for different part types. When the interaction is significant, the reproducibility estimate includes the interaction variance, increasing the reported %GRR. Investigate the source of the interaction before concluding that the measurement system is inadequate.

ndc < 5 with Acceptable %GRR

This occurs when part variation is low relative to measurement variation, even though the measurement variation is low relative to the tolerance. The measurement system is adequate for tolerance-based decisions (pass/fail against specs) but inadequate for SPC (distinguishing between parts within the tolerance). Which evaluation matters depends on the application. For acceptance inspection, use the tolerance-based evaluation. For SPC and continuous improvement, use the study-variation evaluation with an appropriate part sample.

The [Gage R&R calculator](#) reports both the study-variation and tolerance-based evaluations with ndc, allowing the user to assess the measurement system against the appropriate standard.

VII — Related Instruments

INSTRUMENT	SPECIFICATION
Gage R&R Calculator	Crossed and nested Gage R&R. ANOVA method. %GRR, ndc, variance components.
Cpk / Ppk Calculator	Capability analysis after measurement system qualification.

[← All Whitepapers](#)

SVEND is the consulting practice of Eric Wolters – operational excellence, quality management and strategic planning. ILSSI accredited, SCMEP Technical Partnership Provider.
svend.ai/whitepapers/msa-beyond-gage-rr/ · published 1 April 2026