METROLOGY AND MEASUREMENT SYSTEMS
VOL. XIV, NUMBER 4 (2007)

Jozef Petrík, Vojtech Špetuch

Technical University of Košice,
Faculty of Metallurgy, Slovakia
jozef.petrik@tuke.sk

# THE DEPENDENCE OF BRASS HARDNESS MEASUREMENT SYSTEM ON TIME

Measurement system analysis (MSA) was used for estimation of hardness (HBS) measurement system capability. The time dependence of capability was evaluated. Experimental material Cu-Zn-Al cast alloy was used as shape memory alloy. The average and range method with 2 appraisers, 10 samples and 3 trials (indentations). was used to quantify repeatability (% EV), reproducibility (% AV) and part variation (% PV – variation between samples). The measurement was repeated 5 times in intervals of 7 days. As the values of %GRR are between 89.4% and 63.7% and values of "ndc" between 0.71 and 1.71 the used hardness measurement system is not capable and it depends heavily on the long-term variability.

Keywords: time dependence, % GRR method, shape memory alloy, hardness HBS

## 1. INTRODUCTION

When you manufacture products, you want to monitor the output of your machines to make sure that they are products that meet the customer's specifications. All manufacturing organizations need to know the quality of the measurement data used in controlling and monitoring processes. One of the most common reasons for low quality data is too much variation in the data. Much of the variation in a set of measurements is due to the interaction between the measurement system and the environment. If the interaction generates too much variation, then the quality of the data may be so low that the data are not useful. The variability of the measurement system must be small compared with the variability of the manufacturing process.

The overall variability can be separated into three areas: samples (products or samples – differences between individual manufactured pieces), appraisers or operators (reproducibility – can two different people get the same measurement using the same measuring instrument or gage?) and equipment (repeatability – can the same appraiser perform the same measurement using the same gage on the same product in two or more trials?).

Measurement system analysis (MSA) is an experimental and mathematical method of determining how much the variation within the measurement process contributes to overall process variability. MSA involves GRR (gage repeatability and reproducibility – Gage R&R) studies to evaluate measurement systems. It is a system designed to help engineers and quality professionals assess, monitor, and reduce measurement system variation. It teaches how to conduct measurement system studies including linearity, stability, repeatability and reproducibility. MSA helps conform to ISO 9000 and ISO/TS 16 949:2002 requirements as well as AIAG standards. Measurement Systems Analysis teaches how to develop working standards that are traceable to the ISO.

The oldest of the hardness test methods in common use today, the Brinell test is frequently used to determine the hardness of forgings and castings that have a grain structure too course for Rockwell or Vickers testing. By varying the test force and ball size, nearly all metals can be tested using a Brinell test. The Brinell hardness number is calculated by dividing the load applied by the surface area of the indentation.

## 2. MATERIAL AND EXPERIMENTAL PROCEDURE

Repeated measurements of hardness [1, 2] process capability indicated time variability of capability indices. The aim of the submitted work is to study the influence of long term variability – repeated measurement (week-to-week) on the capability of measurement system analysis of Brinell hardness of Cu-Zn-Al cast brass – the shape memory alloy. Aluminium has a great influence on its tensile strength. A characteristic feature of brasses containing aluminium is the presence of a superficial film of aluminium oxide, which renders soldering and casting more difficult. A shape memory alloy (SMA, also known as memory metal or smart wire) is a metal that remembers its geometry. After it is deformed, it regains its original geometry by itself during heating (one-way effect) or, at higher ambient temperatures, simply during unloading (pseudo-elasticity). These extraordinary properties are due to a temperature-dependent martensitic phase transformation from a low-symmetry to a highly symmetric crystallographic structure. The three main types of SMA are the Cu-Zn-Al, Cu-Al-Ni and Ni-Ti alloys. Cu-Zn-Al shape memory alloys are a promising alternative to others since they are comparatively much cheaper and easier to produce and process.

The material for investigation was delivered as $\phi$ 40 mm × 20 mm cylinders. The average Zn content was 12.55 wt.% and that of Al was 8.52 wt.%. The Perkin Elmer 306A spectrometer was used for analysis. The surface measurement was prepared in a standard way by grinding through a series of gradually finer silicon carbide water cooled papers. The sequence was 220, 240… and 3000 grit (ANSI/CAMI grit size scale). Finally, it was mechanically polished with $Al_2O_3$, moistened with water and cleaned with ultrasonic cleaning equipment. The polished surface was etched with an agent containing 10 g $FeCl_3$, 30 cm$^3$ HCl and 70 cm$^3$ $CH_3OH$. The average grain diameter was 0.997 mm. Figure 1 describes the typical microstructure of the Cu-Zn-Al shape memory alloy.


Fig. 1. The microstructure of the Cu-Zn-Al brass.

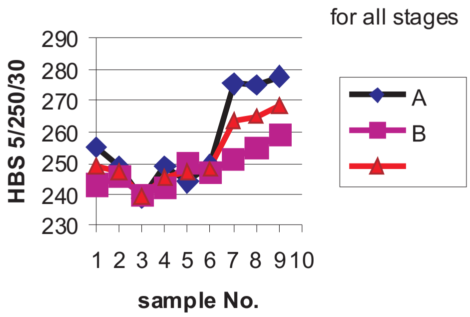


Fig. 2. The average hardness of samples.

The universal hardness tester HPO 250 in HB variation was used as the measuring equipment. The hardness was measured according to standard STN ISO 6506 with load 2.452 kN, ball diameter 5 mm and load time 30 s. For maximal permissible error and repeatability, the tester was up to standard STN ISO 6506-2.

The number of samples (parts) and trials (repeated measurements) depends upon the significance of the characteristic being measured and upon the confidence level required in the estimate of measurement system variation. As with any statistical technique, the larger the sample size, the less the sampling variation and the resultant risk will be present. As a rule, 10 samples, 3 trials (repeated measurements on each sample) and 2 appraisers are used for tests. If possible, the appraisers who normally use the measurement equipment should be included in the study [3].

Discrimination (readibility or resolution) is the minimum amount of change from a reference value that an instrument can detect and faithfully indicate. The measure of this ability is typically the value of the smallest graduation on the scale of the equipment's measurement system. A general rule of thumb is that the measuring instrument discrimination ought to be at least one tenth the process variation. If the measurement system lacks discrimination (effective resolution), it may be an inappropriate system to identify the process variation. This requirement was satisfied, because the value of the smallest tester graduation was 0.91 HBS 5/250/30 in average and average standard deviation (SD) for all hardness values was 18.35 HBS 5/250/30.

The measurement was carried out by two (A and B) skilled appraisers. Each of them carried out 3 trials on each sample. The measurements were made in a random order to ensure that any drift or changes that could occur will be spread randomly throughout the measurement. The measured values are in Figs. 2, 3 and 4. The difference between the hardness of all samples centre and their external surface is considered to be not statistically significant according to results from an unpaired $t$ test to compare two means.

Grubbs' test (with significance level $\alpha = 0.05\%$) detected no outliers. Statistical outliers would indicate that the process is suffering from special causes and is out of statistical control. Ideally, the causes of outliers should be eliminated and new data will be obtained [4].

The standard methods of MSA assume normal probability distribution. In fact, there are measurement systems that are not normally distributed. When this happens and normality is assumed, the MSA method may overestimate the measurement system error. Therefore, before use, the data should be checked to confirm that their distribution is approximately normal. The most simple check is probability plotting, which gives indications of unusual and non-normal distributions.

The normality was evaluated by a normal probability plot, using software Freeware Process Capability Calculator by Symphony Technologies. The normality of all samples, measured by particular appraisers was confirmed.

The quality – capability of the analyzed process is defined by the statistical properties of multiple measurements obtained from a measurement system operating under stable conditions. A measurement system with a large variation may not be appropriate for use in the analysis of a manufacturing process because the variation of the measurement system may mask the variation in the manufacturing process.

The GRR method – combined estimate of measurement system repeatability and reproducibility, described in [3] with confidence 99% and coverage 99% (5.15 $\sigma$) was used for capability evaluation. Periodic GRR studies make it easy to establish and monitor the performance of the equipment. A GRR study can quickly establish the short-time performance of the equipment, including appraiser influence. The method will allow the measurement system's vaanriation to be decomposed into two separate components, reproducibility and repeatability, but cannot express their interaction.

The %GRR value, determining the process capability, partial indices %EV, %AV and %PV were evaluated as well. For evaluating the long-term variability of capability indices, the hardness tests were repeated 5 times in weekly intervals (stages).

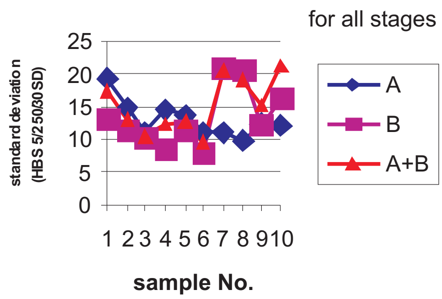


Fig. 3. The SD of samples hardness.

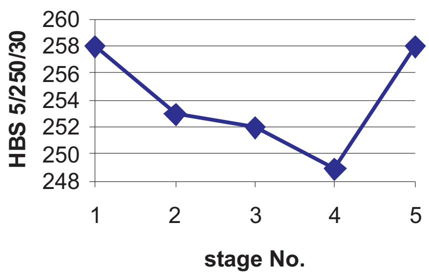


Fig. 4. The average hardness of stages.

## 3. RESULTS

The measurement system ought to be under statistical control before capability is assessed. This means that under repeatable conditions, the variation in the measurement system should be due to common (stochastic) causes only, not due to special (determinable) causes.

The range control chart is used to determine whether the process is under control. If all ranges are in control, all appraisers are doing the same job. If one appraiser is out of control, his method differs from the others.

If all appraisers have some values out of control ranges, the measurement system is sensitive to appraiser technique and needs improvement to obtain useful data. With respect to Table 1 (%R), the condition of system statistical control was not fulfilled at the 3$^{rd}$ (one point, measured by appraiser B) and the 5$^{th}$ (one point measured by appraiser A) stages.

The area within the control limits represents the measurement sensitivity ("noise"). Since the sample group used in the study represents the process variation, approximately one half or more of the averages should fall outside the control limits. If the data show this pattern, then the measurement system should be adequate to detect part-to-part variation and the measurement system can provide useful information for analyzing and controlling the process. If less than a half of them falls outside the control limits then either the measurement system lacks adequate effective resolution or the sample does not represent the expected process variation. With respect to Table 1 (%X, values outside of control limit), this condition was not satisfied.

Table 1. Statistical control of measurement system.

| stage | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| %R | 0 | 0 | 5 (B) | 0 | 5 (A) |
| %X | 10 | 15 | 40 | 35 | 40 |

The criterion whether the measurement system capability is satisfactory depends on the rate of the manufacturing production process variability that is "consumer" by measurement system variation. The %GRR index refers to the contribution of the measuring instrument to the variability. Its value represents the process capability. %GRR < 10 % is generally considered to be an acceptable measurement system. This translates directly into lower uncertainties. %GRR > 30 % is considered to be not acceptable – every effort should be made to improve the measurement system.

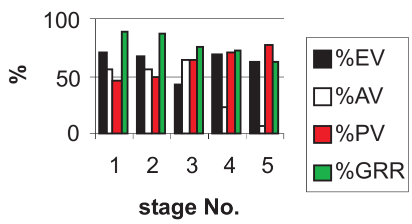


Fig. 5. The capability indices at individual stages.

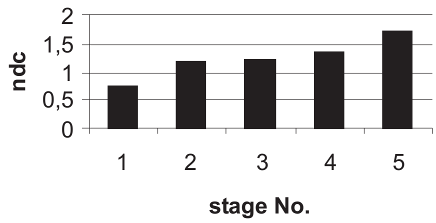


Fig. 6. The value of "ndc" at individual stages.

The analyzed system is not capable because its %GGR value is between 63.7% and 89.4% at individual stages with an increase of capability towards the $5^{th}$ stage. It is possible, that non – capability is typical for hardness, but also for microhardness measurement [5, 6]. However, it is difficult to achieve only 10% variation in hardness testers. The typical results of dead-weight testers are between 15 to 20%. Older testers in poor condition give much worse results [7].

Consistency is the difference in the variation of measurements taken over time. It may be viewed as repeatability over time. Repeatability is the inherent variation or capability of the equipment itself. Repeatability is commonly referred to as equipment variation (EV), although this is misleading. In fact, repeatability is common cause (random error) variation from successive trials under defined conditions of measurement.

Possible causes for repeatability are equipment, standard, method, appraisers lack of experience, environment, wrong gage for the application. In consideration of measurement according to a standard method and standard measurement environment, the %EV value is the foremost evaluation of hardness tester properties. With respect to Fig. 5, the metrological characteristics of the tester, specified by %EV, are steady without a trend to drift.

The %AV index represents the influence of appraisers on variability, for example their competence, perceptions, skills, disciplines and vigilance. It is a function of average values for individual appraisers. Figure 5 indicates progressive improvement of appraiser influence.

The %PV index is a function of range of the average hardness of individual samples. It is sensitive to the variability influence between measured samples. Its value indirectly defines the suitability of used equipment for measurement. The value of %PV above 99% is for very accurate equipment, above 90% for suitable, above 70% for satisfactory and above 50% for inaccurate one. Equipment with a value up to 50% is unsuitable. The used tester can be classified as satisfactory, but also unsuitable (at the first stage) for distinguishing differences in hardness between samples.

The number of distinct categories ("ndc", based on Wheeler's discrimination ratio) of the measurement systems divides a process into "should be greater than" or "equal to 5", but values between 2–5 may be conditionally used for rough calculations. Because, as Fig. 6 indicates, the values of "ndc" are lower than 2 (two), the measurement system is of no value in controlling the process. It is all noise and one sample cannot be said to be different from another.

The charts with sample size 1 were used for evaluation of statistical control of hardness. There are times when, as in this case, it is inappropriate to consider samples with size greater than 1. In this case the ranges are calculated by taking the difference between one value and the next. Effectively, each sample size for range is 2. The plotted values are the individual values of hardness, the central line (CL) is the mean of hardness of all stages, appraisers and samples. Note that in this particular chart the range values are not independent. With respect to Fig. 7, from the long-term viewpoint, the hardness measurement system for this measurement is stable.

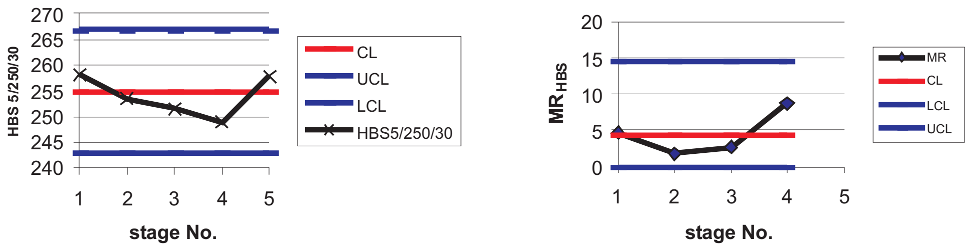


Fig. 7. The control charts of hardness.

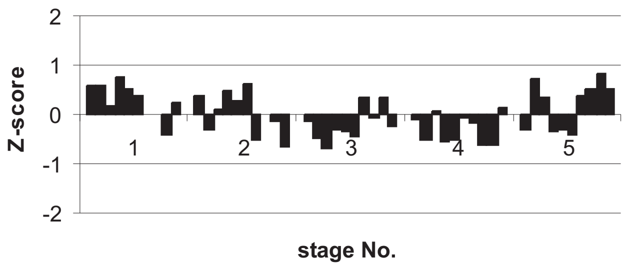


Fig. 8. Z-score.

The Z-score method, routinely applied in interlaboratory comparison tests, was used for validation of the results mentioned above. One stage is equal to one laboratory in standard appliance. The value for an individual sample is:

$$z_i = \frac{x_i - \bar{x}}{s}, \tag{1}$$

where: $x_i$ is the average hardness on one sample at one stage, $\bar{x}$ the average hardness on one sample at all stages and "s" is the standard deviation for one sample at all stages. The results $|z_i| \leq 2$ are satisfactory and $|z_i| \geq 3$ are unsatisfactory [8]. As can be seen in Fig. 7, the results are satisfactory for all stages.

## 4. CONCLUSIONS

1. The measurement process of brass hardness is not capable.
2. Analysis indicates that the measurement process capability is stable since there are no obvious special causes effects visible.
3. Repeated measurements lead to the improvement of capability.
4. The hardness tester is stable, without significant drift.

### ACKNOWLEDGEMENTS

This work was supported by the Slovak Grant Agency for Science VEGA 1/4141/07.

### REFERENCES

1. Petrík J., Špeťuch V.: *The dependence of hardness measurement system capability on the time*. In: Transfer 2005, pp. 429–432. (in Slovak)
2. Petrík J., Špeťuch V.: *The Dependence of Hardness Measurement System Capability on the Time*. Transactions of the Universities of Košice, no. 3, 2005, pp. 27–33.
3. Measurement system s analysis (MSA). Reference manual. Third edition. 2003, pp. 102–120.
4. Betteley G., et al.: *Using statistics in industry*. Prentice Hall International. Hemel Hampstead, 1994, pp. 186.
5. Petrík J., Palfy P.: *Quality of process of measuring microhardness*. In: Jakost – Quality 2006. Ostrava, pp. 14. (in Slovak)
6. Petrík J., Špeťuch V.: *The quality of hardness measurement process*. Materials engineering – Materiálové inžinierstvo, vol. 14, no. 1, 2007, pp. 37–41. (in Slovak)
7. Tobolski E.: *Uncertainty in Hardness testing.* (part 2). Advanced materials & processes, vol. 161, no. 5, 2003, pp. 25.
8. *Methodical guidelines for accreditation. Proficiency testing schemes. MSA 0113-98*. Bratislava: Slovenská národná akreditačná služba, 1998. pp. 32–34. (in Slovak)