Statistical Characterization and Control of Variation in the Manufacture of Standard Test Blocks used for Rockwell Hardness Testing

by

Hans J. Laudon
B.S. Mechanical Engineering, Rochester Institute of Technology (1990)

Submitted to
the Department of Mechanical Engineering and the Sloan School of Management
in Partial Fulfillment of the Requirements for the Degrees of

MASTER OF SCIENCE IN MECHANICAL ENGINEERING
and
MASTER OF SCIENCE IN MANAGEMENT

at the
Massachusetts Institute of Technology
June 1996

Signature of Author ....
Hans J. Laudon
May 10, 1996

Certified by ....
Roy E. Welsch, Thesis Supervisor, Professor of Statistics and Management Science

Certified by ....
Kenneth C. Russell, Thesis Supervisor, Professor of Metallurgy, Professor of Nuclear Engineering

Certified by ....
David E. Hardt, M.E. Reader, Professor of Mechanical Engineering

Accepted by ....
Jeff A. Barks, Associate Dean, Sloan Master's and Bachelor's Programs

Accepted by ....
Ain A. Sonin, Chairman, Department of Committee on Graduate Students

Statistical Characterization and Control of Variation in the Manufacture of Standard Test Blocks used for Rockwell Hardness Testing

by

Hans J. Laudon

Submitted to
the Department of Mechanical Engineering and the Sloan School of Management
in Partial Fulfillment of the Requirements for the Degrees of
Master of Science in Mechanical Engineering and
Master of Science in Management

ABSTRACT

The uniformity variation of hardness across a test block, the reference material for calibrating a Rockwell tester, is masked by the variation of the commercial tester system used to measure the block. The sources of total measurement variation can be isolated by means of a statistical model introduced in this thesis. Measurement feedback from a precision deadweight tester, as maintained and operated by NIST[1], serves in the development of these models; it also plays a key role in achieving an effective process control strategy for the manufacture of hardness test blocks.

The author develops the Calibration Capability index, Cc, to characterize the interaction between variation sources in the stages of (manufacturer) block calibration and (customer) tester validation. Along with the statistical process metrics of Process Capability, Cpk, and Range Capability, C_R, the performance of the test block manufacturing process can be adequately measured with respect to customer requirements for the reference standard system. These metrics provide guidance toward the future product and process improvements and may be applied in organizational goal-setting.

The application of statistical process control (SPC) methods using conventional Shewhart Xbar, R and s control charts is shown to be practically feasible and beneficial in controlling the uniformity of hardness test blocks. A methodology that cross-references a deadweight tester system distinguishes the causes of 'out-of-control' conditions between either the measurement system or the block process.

In order to benefit from SPC methods, management must tailor a production environment that fosters problem-solving with communication along the entire process chain and that motivates conformance to standardized procedures. Management methods for SPC introduction and related organizational issues are presented.

Thesis Advisors: Roy E. Welsch, Professor of Statistics and Management Science
Kenneth C. Russell, Professor of Metallurgy, Professor of Nuclear Engineering

[1] National Institute of Standards and Technology, United States Commerce Department, Gathersburg, MD

ACKNOWLEDGMENTS

I gratefully acknowledge the support and resources made available through the Leaders for Manufacturing (LFM) Program, a partnership between MIT and U.S. manufacturing companies.

I would also like to express my appreciation to the Instron Corporation for their commitment and support in providing the learning laboratory for this thesis. The efforts and attention by my supervisors, Thomas Zarrella and Edward Tobolski, are gratefully recognized. Thanks are also extended to the individual members of the Uniform Block Excellence Team (UBET) for their hard work and courage. The review by Ken Ryan was very appreciated.

The counsel and endurance in review provided by my research advisors, Roy Welsch and Ken Russell, are acknowledged. I would also like to thank Dave Hardt for setting up the LFM internship and for his inspiring process control course.

I am grateful for the contributions from the National Institute of Standards and Technology in Gathersburg, MD. In particular, Sam Low and Dave Pitchure, are recognized for providing the hardness measurement data that was critical to the models of this thesis.

I would also like to use this venue to express my deepest gratitude to all the people that made my graduate education at MIT possible, including Bill Hutchings, Ron Salzman, and Horst Engelbrecht of LIGHTNIN; Mark Kempski of Mechanical Engineering at RIT; and Raymond Drago, who initiated my aspiration for LFM.

Special thanks to my mother and sister for their life-long encouragement and long-distance friendship.

These words cannot express my appreciation to Penny Lewis who provided both emotional and financial support and was a true companion during this challenging time.

TABLE OF CONTENTS

Part II: CAPABILITY METRICS FOR PRODUCT AND PROCESS PERFORMANCE

Part III: STATISTICAL METHODS FOR PROCESS CONTROL

Part IV: ORGANIZATIONAL AND MANAGEMENT ISSUES

Chapter 1 Introduction

Until the present time, the unofficial reference standard in the U.S. for Rockwell hardness was developed and maintained by the leading commercial manufacturers of Rockwell hardness testing systems. Wilson Instruments, a division of Instron Corporation, has been at the forefront of technology development for commercial Rockwell testing systems for over 75 years.

The hardness standard for Rockwell hardness testing is embodied in the form of a flat, disk-shaped block of reference material, called a test block. The material composition and states are specifically controlled to yield a particular hardness level within the available Rockwell scales. Instron manufactures the test blocks and calibrates them to their inherent hardness level in the Wilson Standards Laboratory using a bank of six 'standards' testers. The calibrated hardness is engraved on the edge of the test block and is also referenced in a calibration certificate supplied with each test block. Refer to the calibration certificate of Appendix D.

The new Wilson 600 Series Rockwell tester was integrated into the laboratory early in 1996. Refer to Appendix A. This tester technology, which replaced the 500 Series tester, represented the commercial state of the art in the Rockwell testing industry available at the time.

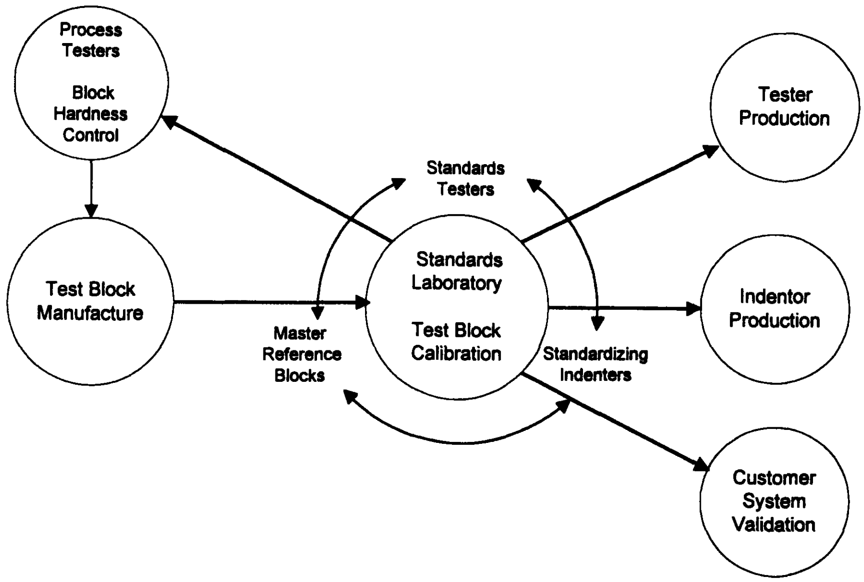
The test blocks, calibrated to the Wilson Standard, are in turn used to calibrate the nominal machine settings for production Rockwell hardness testers supplied to customers. Test blocks are also applied in the set up the Rockwell testers that gage the acceptable accuracy of indenters in their manufacture. A closed loop system of standard traceability thus exists in which the manufacturer must control its reference master test blocks, indenters and standardizing testers in order to prevent the presence of variation from causing a transient drift in the nominal hardness levels of the reference standard.

Next to satisfying the internal production needs for a reference standard, Instron markets calibrated test blocks to users of commercial Rockwell hardness testers, the majority for use with machines supplied by Instron/Wilson. The purpose of the test block is to allow the customer to verify the calibration set points of their hardness testing system which over time is subject to mechanical wear and usage drift. This calibration verification process is termed 'validation'.

If the tester should fail the validation, a diagnosis of the equipment condition results in a series of minor corrective actions as outlined in the user manual. Only once these corrective actions have been exhausted, is a machine calibration adjustment, by a qualified Wilson service technician performed.

Refer to Figure 1-1 for a schematic of how the Instron hardness reference standard is maintained in the physical form of hardness test blocks.

Figure 1-1: Instron Reference Standard System through Hardness Test Blocks



1.1 The Rockwell Hardness Test

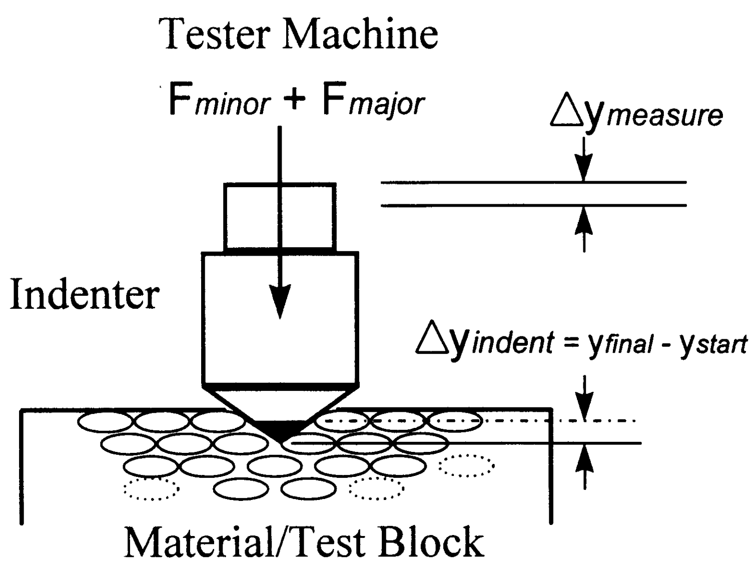
The Rockwell hardness test consists of a mechanical measurement for the relative displacement of a diamond-tipped indenter (or ball penetrator) as it penetrates the surface of a test specimen between successive minor and major loads [5]. Refer to Fig. 1-2. The 10 kgf minor load seats the indenter at its starting reference (SET) position. The Rockwell scale (e.g. C, B, A) is determined by the selection of major load. For the Rockwell 'C' scale of this study, the major load is 150 kgf. Following the application of the major load, the Rockwell tester switches back to the minor load to allow elastic recovery at its final position. The resulting hardness number, which is read from a digital display of the tester, represents a secondary linear calculation performed by the tester logic, as follows:

$$HR = 100 - [(y_{final} - y_{start})/2\ \mu m] \qquad \text{(for a diamond-tipped indenter)}$$

Each Rockwell number constitutes a material penetration of 2 μm (8×10^{-5} in.). An infinitely hard material thus has a Rockwell hardness of 100. The reader will note that the relevant measurement increment is smaller than the grain sizes of most fine-grained heat treated steels of mean grain diameters ranging from 8 μm (ASTM/ISO G 11) to 22 μm (ASTM/ISO G 8) [20].

The Rockwell hardness test is destructive by nature as it leaves a small indentation in the surface layer of the test material. As the material location cannot be re-tested, no two sequential tester operations sample the same group of material grains. Due to the restricted size of the standard measurement area of 4.0 in.^{2} and the requirements for minimum spacing between indents, a finite number of hardness measurements can be performed on the test block [4]. Depending on hardness (indent diameter) the maximum feasible indent quantities range from approximately 120 to 150 indents per block depending on hardness [3].

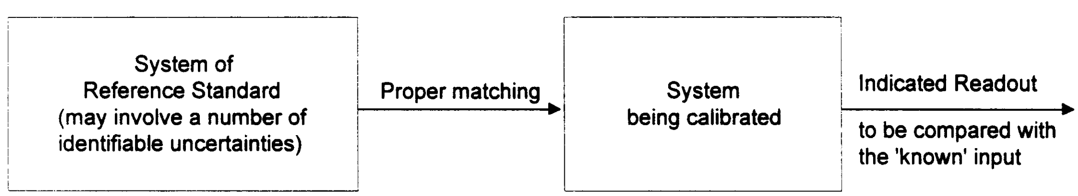
Figure 1-2: The Rockwell hardness test



1.2 Existing Models for Reference Standards of Measurement Devices

The generic cross-referencing to a known manufacturer's standard for a commercial measurement device is depicted in Figure 1-3 below:

Figure 1-3: Block diagram showing a generic calibration procedure [1]



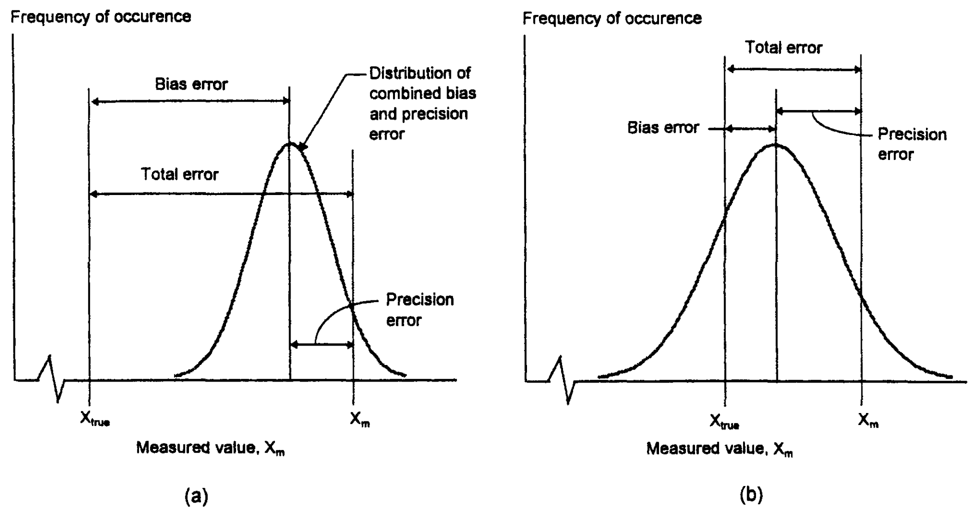
The uncertainty of the reference standard can be described as a combination of [1]:

- bias or accuracy error on the nominal calibration hardness, the average of n hardness readings
- precision error due to the variation of individual hardness readings

Refer to Figure 1-4 for a graphical representation of these error types.

A generic rule of thumb exists for measurement devices that the uncertainty of the standard system should be no more than one-tenth of the system being calibrated [1]. As demonstrated in the following chapter, the current state of technology in the Rockwell testing industry does not allow this heuristic to be satisfied.

Figure 1-4: Bias and Precision Errors: (a) bias error larger than the typical precision error, (b) typical precision error larger than bias error [1]

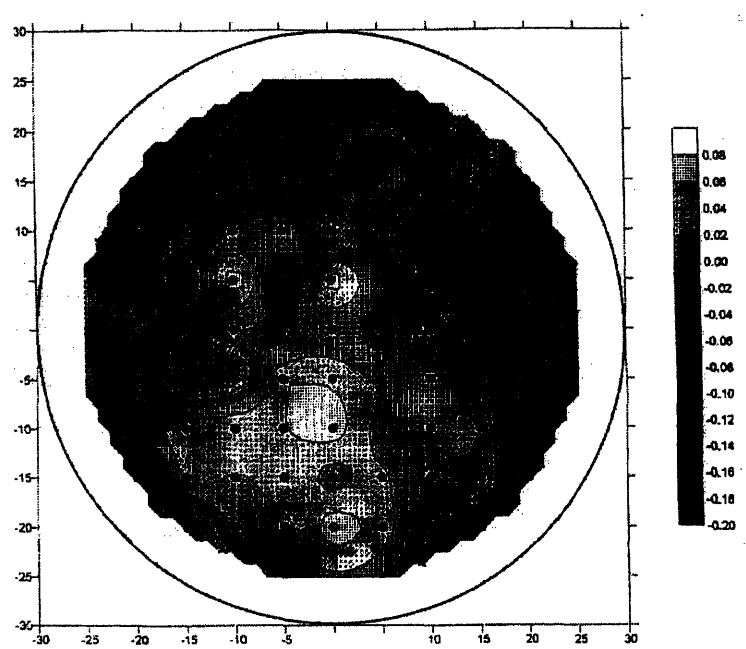


1.3 Hardness (non-)uniformity of a Test Block

A constituent of the measurement precision error is the variation of hardness across the measurement surface of the test block, termed *non-uniformity* [4]. Refer to Figure 1-5 for a graphical representation of a possible irregular hardness distribution for a test block. The non-uniformity of the block is treated as a material state that is to be determined by measurement.

It is thus the objective of Instron to minimize the block hardness non-uniformity in the manufacture of the test block.

Figure 1-5: Representation of Hardness Variation across the Surface of a Test Block



1.4 A New National Rockwell Hardness Standard

In 1994/95 the National Institute of Standards and Technology (NIST) of the U.S. Commerce Department embarked on establishing a U.S. national standard for Rockwell hardness, independent of those standards maintained by the respective equipment manufacturers. Rockwell hardness reference standards in several other countries are similarly defined by a designated government agency, for example JBS in Japan, DIN in Germany or IMGC in Italy.

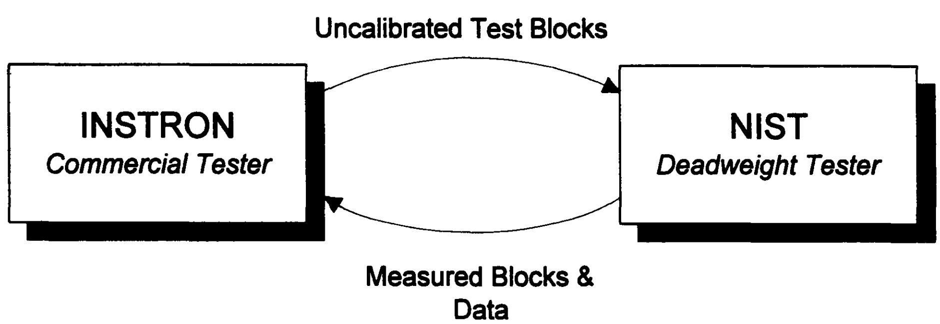
The national standard will be maintained by NIST in Gathersburg Maryland, through strict control of a specially designed deadweight tester. The key characteristics of the deadweight tester are direct load application by a referenced mass, as well as low total measurement variation.

During his internship in the summer and fall of 1996, the author helped Instron develop an improved process for a steel test block of high uniformity for the supply of high quality test blocks to NIST. The blocks covered the hardness range of 25 to 63 on the Rockwell 'C' scale. This special block, termed the 'Large' block, was thicker (0.63 in. vs. 0.38 in.) and had a larger diameter than the prior block geometry of 'Regular' blocks sold by Instron. Refer to Appendix B.

Following an intercomparison study of candidate suppliers, NIST awarded their purchase of uncalibrated test blocks to Instron Corporation on the basis of particularly tight specification requirements for test block hardness precision or uniformity. NIST calibrates these blocks using their high precision deadweight tester to create Standard Reference Material™, referenced to the U.S. national hardness standard. NIST determined the acceptable uniformity from the total measurement range of a large number of distributed indentations (n = 25 to 77) using their deadweight Rockwell tester.

In order for Instron to confirm the acceptable quality of their test blocks supplied to NIST, the measurement data of block qualification from the deadweight tester was provided by NIST. A set of Large grade blocks, which were measured using the deadweight tester at nominal hardnesses of HRC 30, 40, 50 and 60, were also returned.

Figure 1-6: Information Feedback from National Inst. of Standards and Technology (NIST)



1.5 Standard Specifications for Rockwell Hardness Testing

Specification standard ASTM E-18 defines the Rockwell testing methods and the corresponding specification limits for the equipment, to which the all manufacturers of Rockwell hardness systems comply. The standard is established and regularly revised by a subcommittee of ASTM that is largely comprised of the industry players in hardness testing, both manufacturers and users, including Instron.

Allowable tolerances for test block accuracy at each nominal hardness level are specified in ASTM E-18. Refer to Table 1.1 below . Instron applies this maximum standard tolerance to the average hardness calibrated for the block using n = 6 measurements. The standard tolerance is provided on the calibration certificate supplied with every test block and is engraved on the edge of each test block. The customer uses these tolerances as confirmation boundaries when measuring the hardness of the same block using his/her testing system using maximum of n = 5 measurements in validation.

In addition, the precision error or degree of measurement variation is limited by the specification. The standard refers to 'maximum non-uniformity of standardized test blocks'; these tolerances are applied to the hardness readings, which are subject to measurement noise in addition to test block non-uniformity. Non-uniformity of hardness for standardized test blocks is controlled by a maximum allowable range for n = 5 randomly-placed measurements. The relevant 'non-uniformity' specifications for the Rockwell 'C' scale of this study along with the standard tolerance values for the calibration hardness are given in Table 1.1 below:

Table 1.1: Standard specification of ASTM E-18 for test block measurement variation [4]

Rockwell 'C' Scale Nominal Hardness	Allowable Range for n=5 measurements ASTM E-18, Table 22	Maximum tolerance on the calibrated average hardness ASTM E-18, Table 21
HRC 60 and greater	0.5	± 0.5
Below HRC 60	1.0	± 1.0

It should be noted that ASTM E-18 does not specify the type or condition of the Rockwell testing system by which the non-uniformity is measured. Clearly, the allowable range that serves to bound the precision error, includes components of variation attributable to the response of the particular testing system.

The measurement and calibration methods of ASTM E-18 take into account the practicality of industrial uses and the costs of expending test blocks. As a result, the metrics for variation are simplified using *Ranges*. In addition, the variation is qualified using a low sample size of n=5 measurements to ensure block longevity.

The Japanese Rockwell standard JIS B 7730 [8] and the German standards DIN 51303 [6] and 51304 [7] also specify allowable ranges and mean tolerances. However, the application of statistical methods is found in the NAMAS hardness standard document NIS 0406, in which the variation of block and standards tester system is characterized using standard deviations (and variances). NIS 0406 also provides a methodology by which to compute 95% confidence intervals for the calibrated average hardness using the t-statistic for small sample sizes [9].

1.6 Measuring System Variability: Gage Repeatability and Reproducibility

Instron attempts to characterize the variation of the Rockwell measuring instrument using the gage repeatability and reproducibility study (GRR), as applied to other conventional measuring devices e.g. micrometers. The set of designed experiments varies operators and trials over a larger sample of 10 different test blocks. The GRR study yields an assessment of how much of the 'process tolerance' is used up by the variation of the Rockwell test system (repeatability or within operator/device variability) and variation among operators (reproducibility or between operator variability) [2].

The shortcomings of this methodology when applied to Rockwell testing are:

- the inherent block hardness variation is not accounted for [2]. Unlike other measuring devices the same locations of constant properties cannot be remeasured.
- it converts average measurement ranges, stemming from small sample sizes (n=3), to estimates of sample variances using conversion factors. The conversion factors are based on expected values for the assumption of a normal distribution of individual measurements. Thus, the calculated sample variances used are subject to sampling error and estimation error.
- the resulting % R&R metric is dependent on a process tolerance subject to selection. In industry applications, the tolerance reflects process specification limits for a particular part's hardness; 5 to 6 Rockwell points is typical.
- the hardness variation is dependent on the hardness level, which is often not specified.
- the GRR procedure is sufficiently complex that it cannot be easily and frequently repeated as an on-going performance metric.

As a result, it is difficult for the user (or even a trained statistician) to render any rooted meaning from the results of the GRR study. In addition, a GRR evaluation applied to the process of test block manufacture is subject to much tighter process specification limits, resulting in large % R&R's (>>10%). The GRR study is also not conducted often enough to understand the influence of different types of blocks, hardnesses or tester conditions.

The Calibration Capability Index, Cc, is developed as an improved alternative to the GRR [Refer to Chapter 5].

1.7 Problem Statement

In order to control the manufacture of a test block of low non-uniformity, the manufacturer must be able to confirm its changes in state by measurement. Instron does not currently employ statistical process control methods in the manufacture of hardness test blocks or for the control of its standardizing testers used to measure the process blocks.

The measurement noise of the commercial hardness testers employed by Instron is sufficiently large and potentially dynamic, such that the contribution of block non-uniformity to the total measurement variation cannot be isolated. The very nature of a reference material of best-possible quality prescribes that its manufactured variation is straining the limits of measurement capability. This is particularly the case if a tester technology similar to that used by the customer base is applied in test block production. The existing analytical methods of quantifying variation through measurement ranges of small sample sizes have not sufficed.

As a result, improved methods for characterizing the sources and the nature of variation must be developed in order to isolate and control the variation for test blocks of improved uniformity, such as the Large grade blocks manufactured for NIST.

In their literature study regarding the state of knowledge on hardness test blocks, OIML (Bureau International de Metrologie Legale) cites the challenges in controlling block uniformity with a hint at a path toward a solution : "One could reduce the sampling uncertainty by increasing the number of indentations but the cost of calibration would increase and the value of the block is reduced, as the usable surface is reduced. Consequently, the only reasonable solution is to find a perfect cooperation and interaction between production and calibration of the blocks. To find ways and means to detect production deficiencies from small number of hardness tests. Sampling plans, control charts and other methods of quality control should be employed" [3].

1.8 Thesis Statement

The exchange of measurement data from NIST's deadweight Rockwell tester can be leveraged to characterize the relative sources of measurement variation using tools of applied statistics. Statistical capability metrics can be developed from these models in order to better measure product and process performance in a manner that accounts for the interaction of the variation sources and in a way that captures the customer's validation needs. Statistical process control (SPC) methods can be tailored to meet the challenges of measurement noise and small sample sizes in controlling test block hardness uniformity. These control strategies can be consistent with practical constraints, such that process control can feasibly serve as a foundation for future process and product performance improvement. Organizational competencies and infrastructures must be evaluated in order to assess the overall costs and benefits of SPC implementation for test block manufacture. The author aims to support such planning and evaluation with this thesis.

1.9 Project Scope and Limitations

This study demonstrates methods for statistical analysis and process control using representative sample data collected in the fall of 1995 and January of 1996. A large portion of the data was derived from the process experiments of the *UBET*[2] optimization program for the development of Large grade test blocks. Some of the statistical relationships contained in this study have therefore only been confirmed for steel blocks measured on the Rockwell 'C' scale. As the state of process control and capability is subject to change over time, the author cautions that the sample performance metrics contained herein reflect a snapshot in time. The statistical methodologies are outlined in sufficient detail that the analysis can be repeated in order to assess current states from more recent data.

---

[2] The name 'UBET' was selected by the Uniform Block Excellence Team; they know who they are.

1.10 Terminology and Notes on Equipment Parameters

Terminology, that may be particular to this study, is used to express the different types of equipment under focus. These equipment types constitute different parameters in the systems for Rockwell hardness testing. Each equipment parameter is explored to ensure that the statistical methods developed herein can be universally applied to the spectrum of Instron's product offerings and standardizing equipment.

Test Block Process Types: (for Rockwell 'C' scale only)

Large Grade: The Large grade test blocks were the result of a process optimization program conducted during 1995. Process samples were qualified by NIST as high-quality test blocks that exhibit superior uniformity.
Large blocks are thicker than Regular grade blocks and exhibit a larger outer diameter, which is greater than the effective measurement diameter in order to avoid potential edge effects. See Appendix B.

Regular Grade: The Regular grade blocks are derived from a manufacturing process *prior* to the optimization program for the Large blocks. These blocks are thinner and their outer diameter represents the outer bound of the available measurement area. See Appendix B.

Rockwell Tester Systems: 150 kg major load, 10 kg minor load

Instron 500S: Standards Lab tester Model B523R, Serial No. 80195408;
Indenter Serial No. 940191
This represents a standardizing tester, that in January of 1996 was being replaced in the Instron Standards Laboratory by the more current 600 tester technology. Refer to the equipment diagram of Appendix A.

Instron 600S: Standards Lab tester Model A653RMT-4, Serial No. 97328502;
Indenter Serial No. 95621105
The 600S is a new standardizing tester installed in the Instron Standards Laboratory in January 1996. Refer to the diagram of Appendix A.

Instron 600R: Research 600 tester Model 653 C, Serial No. 97331502;
Indenter Serial No. 95621105
The 600R was a tester system used for internal process measurements of Large grade blocks during the UBET optimization program of 1995. Refer to the equipment diagram of Appendix A.

NIST Deadweight: The deadweight tester is maintained and controlled by NIST in Gathersburg, MD, for purposes of establishing a national hardness reference standard; as such it exhibits extremely low measurement noise.

The reader should note that all indenters used for the hardness measurements in this study have been inspected to be of Instron Standards quality. Only one indenter was applied with each tester system, as matched above. All measurements were conducted by a single operator, the author.

1.11 Reader's Guide

Four main parts make up the structure of this thesis.

The first part takes a fundamental approach to defining the sources of Rockwell measurement variation and describes their empirical behavior using statistical techniques. Chapter 2 derives the components of variances model with respect to simplifying assumptions. Chapter 3 uses this model to show how uncertainty is propagated through the reference standard system from the manufacturer to the customer. A global system model captures the interactions that are defined statistically. Chapter 4 discusses the underlying assumptions that allowed the application of particular statistical methods to the models; the basis for the development of the capability indices and the use of SPC control charts are thereby justified.

The second part introduces three capability indices for measuring product and process performance. Chapter 5 includes a rudimentary statistical development of the Calibration Capability index, Cc, which captures the interactions of the global system model. Sections 5.7, 5.8, 5.9 and 5.10 are aimed at supporting the understanding of the Cc index for the reader that is new to statistics. The final sections of Chapter 5 show how the Cc index can be used to make decisions for product and process improvements. Chapter 6 demonstrates the application of the known Process Capability index, Cpk, in gaging the feedback control performance of the test block heat treatment process. Chapter 7 introduces the Range Capability index.

The third part looks at how the manufacturing process for Rockwell test blocks can be controlled using statistical process control (SPC) methods using a feedback control perspective. Chapter 8 discusses the enablers and barriers to the application of SPC. Chapter 9 presents a comprehensive strategy with tactical details for introducing SPC at Instron Corporation for test block manufacture.

The fourth and final part addresses the organizational issues related to the introduction of statistical methods into a firm with little prior SPC background. Section 10.2 discusses why SPC makes sense as an integral part of the firm's competitive business strategy. Chapter 10 also outlines the transformed roles of management, production associates and external players.

The thesis is concluded in Chapter 11 with recommendations for future work.

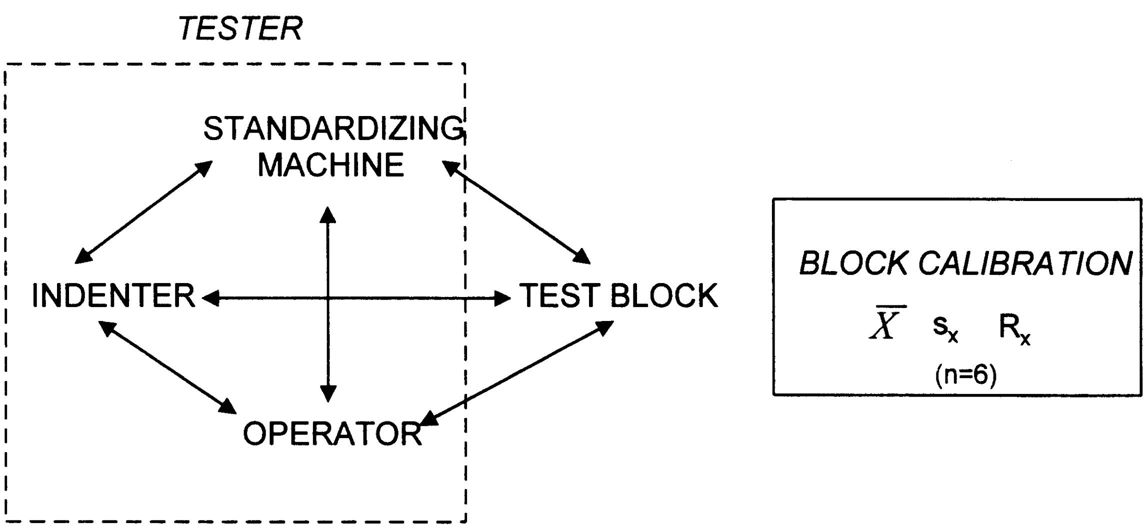
Part I Statistical Characterization of a Rockwell Hardness Standard System

Chapter 2 Modeling and Quantifying the Sources of Rockwell Measurement Variation

2.1 The Lumped Tester Model

Hardness measurements performed on a test block, as any other material, are subject to variability. The sources of this variation are aggregated into the tester machine, the indenter, operator and test block. These sources of total measurement variation in the calibration of a test block are depicted in Fig. 2-1. To date the relative contribution to the total measurement variation from each of these sources has been unquantified by Instron. The variation contribution due to the test block is termed its *non-uniformity* [See Section 1.3].

Figure 2-1: Sources of Measurement Variation for the Calibration of Test Blocks

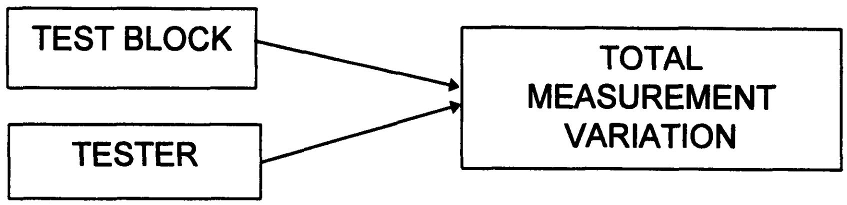


The Gage Repeatability and Reproducibility (GRR) methodology applied by Instron and the hardness testing industry in general attempts to minimize the contributions of block and operator variation in an effort to isolate the nature of the combined machine and indenter variability. Cieplak et al explain how this methodology is largely influenced by inherent non-uniformity of block hardness over the measurement surface [2].

The key to improving tester and indenter technology lies in first being able to quantify and control the hardness non-uniformity of the block. As summarized from the literature study on the subject by OIML: " To assess the true performance of the hardness testing machines (including standardizing equipment as well), it is necessary to take into account the behavior of the test blocks in use, the sampling variability arising from the non-uniformity of the blocks and the environmental changes that may have occurred" [3].

For the subsequent purpose of isolating the variation attributable to the block, the author simplifies the model into two primary sources: the test block and the 'lumped' tester. The 'lumped' tester therefore includes the indenter, operator and tester mechanism sources of variation.

Figure 2-2: The Lumped Tester Model for Sources of Measurement Variation



2.2 Measuring Variation

The author chooses the sample standard deviation, s, and its squared-value, the sample variance, s^2, as the metrics to describe variation. The sample standard deviation is given as:

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} \qquad (1)$$

Though it requires more computational effort, the sample standard deviation is less susceptible to outliers than the sample Range. In addition, the sample standard deviation is a more efficient estimator than Range for the true standard deviation, σ, of normal probability distributions (which will be encountered later) [16].

2.3 General Characteristics of Tester and Measurement Variation

It is hypothesized that the tester variation differs between different types of tester technologies. This is demonstrated by the relative measurement data from hardness measurements performed on four common blocks by three different tester technologies: NIST Deadweight Tester, the Instron 600S tester and the Instron 500S tester.

2.3.1 Relative measurement variation for a set of common blocks

A set of Large grade blocks, covering four hardness levels HRC 30, 40, 50, 60, where measured under controlled conditions by a common operator (the author) using different indenters in each tester. The measurement sample groups are statistically significant with $n \geq 30$; the NIST measurements were conducted with n=68 to 76 indentations.

Figure 2-3 depicts the standard deviations of the measurement samples at each hardness level and Figure 2-4 shows the variance of the same samples.

Since these measurements were conducted on a common block, the comparison of both s and s^2 values at each hardness level allows the conclusion that, in general, the NIST deadweight tester has lower total measurement variation than either the 600S or 500S testers. The 600S also seems to exhibit less variation than the 500S. The conclusion is limited due to the sample size of one block at each hardness level.

2.3.2 Relative Measurement Variation for Equivalent Process Samples

In the course of conducting process optimization studies for Large blocks supplied to NIST, significant measurement data using a controlled 600R tester was collected on blocks that were manufactured in parallel with those delivered to NIST. NIST tested one sample from every four process lots at each hardness level, HRC 25, 45 and 63. Hence, the process samples measured by the Instron 600R tester at n = 25 measurements could be compared to blocks from the same process lots, measured by NIST at n = 25 indentations per block using the deadweight tester. The sample quantity of 25 is deemed sufficient to yield statistically significant estimates of means and standard deviations. Both the Instron and NIST Rockwell measurements were conducted using a common indenter and operator for all of their respective measurements.

The variation of this measurement data given in Appendix C is shown in Figure 2-5 (standard deviation). Figure 2.6 is a plot of the variances based on the average standard deviations shown in Figure 2.5. This graph highlights the trends discussed earlier for the common block measurements.

Two deficiencies in the comparison of NIST to 600R tester measurements are acknowledged:

- the low sample quantity of blocks measured by NIST, 4 vs. 16 to 19 blocks.
- the measurements are not conducted on common blocks

The blocks originated from the same steel material batch; they were heat treated in common lots, and were finished together under a state of process monitoring. Therefore, the author deems the data to be sufficient to provide genuine insights into the relative measurement variation. In addition, the behavior of the data matches that of the common block tests of Figs. 2-3 and 2-4.

The error bounds (max./min.) on the standard deviations on Figure 2-5 demonstrate that the measurement variation of the 600R tester has a rather large spread. The total variability for this tester does not seem consistent from block to block. However, the dramatically decreased spread on the NIST data suggests, though not conclusively, that the 600R tester variation in standard deviation is not attributable to the blocks.

2.3.4 Conclusions on the General Nature of Measurement and Tester Variation

The results depicted in Figures 2.3 through 2.6 support the following conclusions:

- The measurement variation using the NIST deadweight tester is significantly lower than that of the 600 and 500 testers. For a common block, it thus follows that the NIST *tester* variation is substantially lower than that of the 600[3] or 500 testers.
- The variation of 600 measurements decreases with increasing hardness. Figs. 2.3 and 2.4 allow us to extend this response behavior to the 500 tester. At the higher hardnesses of HRC

[3] Note that the measurement results for the 600R and 600S systems are generalized as '600 Tester'.

60 and above, the spread on the 600 variance overlaps into the expected variance of the NIST tester.

- The variation of NIST deadweight measurement is fairly constant with respect to hardness.
- In general, the degree and nature of measurement variation differs between tester technologies and hardness levels.
- Because the measurement variation is a function of hardness for Instron's commercial tester technologies, comparisons and analyses of variation must be performed at each respective nominal hardness levels.

Given these results the question arises: Can we isolate the relative contributions to total measurement variation between the block and tester ? The answer is yes, if we leverage the low variation of the NIST deadweight tester.

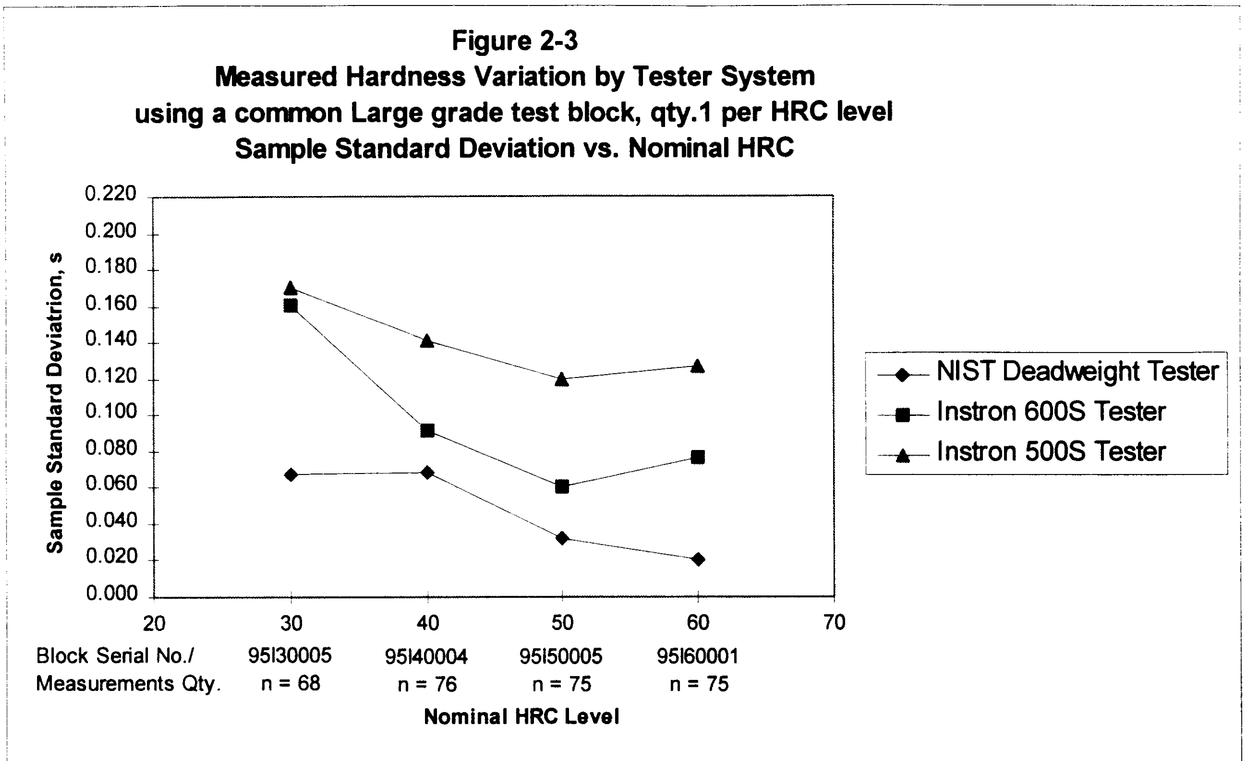


Figure 2-3
Measured Hardness Variation by Tester System
using a common Large grade test block, qty.1 per HRC level
Sample Standard Deviation vs. Nominal HRC
Sample Standard Deviatrion, s
0.220
0.200
0.180
0.160
0.140
0.120
0.100
0.080
0.060
0.040
0.020
0.000
20
30
40
50
60
70
Block Serial No./
Measurements Qty.
95130005
n = 68
95140004
n = 76
95150005
n = 75
95160001
n = 75
Nominal HRC Level
NIST Deadweight Tester
Instron 600S Tester
Instron 500S Tester

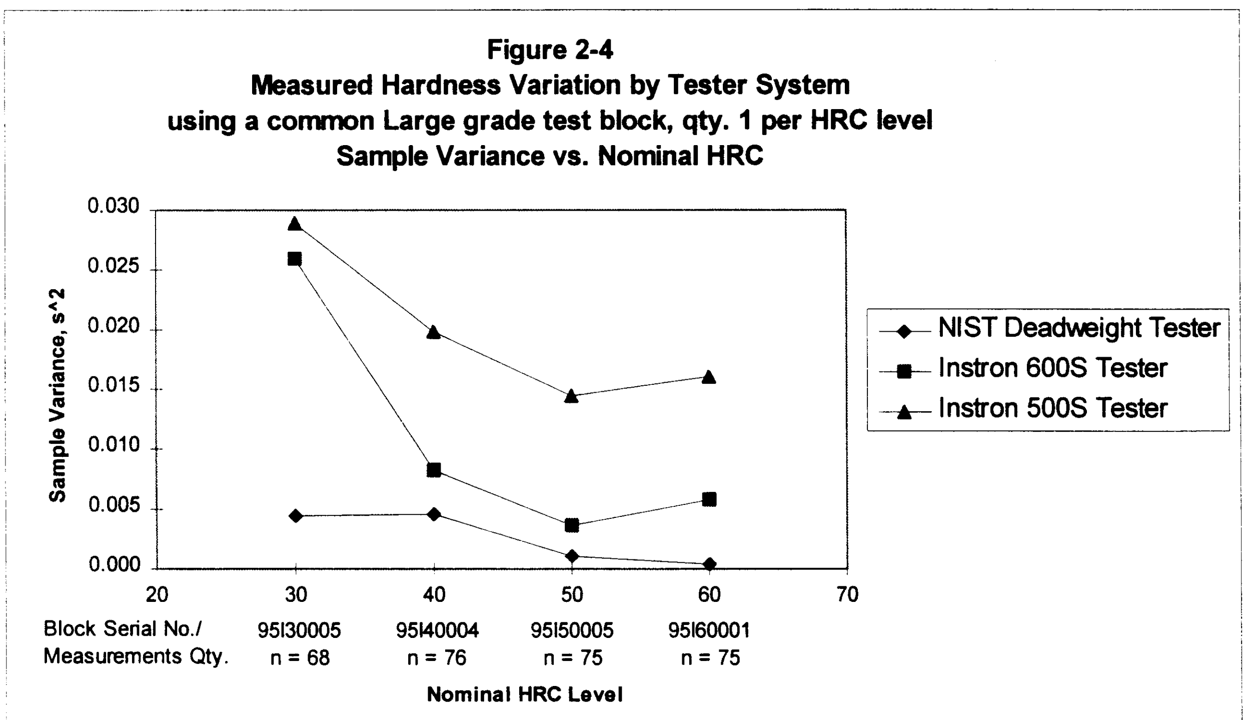


Figure 2-4
Measured Hardness Variation by Tester System
using a common Large grade test block, qty. 1 per HRC level
Sample Variance vs. Nominal HRC
Sample Variance, s^2
0.030
0.025
0.020
0.015
0.010
0.005
0.000
20
30
40
50
60
70
Block Serial No./
Measurements Qty.
95130005
n = 68
95140004
n = 76
95150005
n = 75
95160001
n = 75
Nominal HRC Level
NIST Deadweight Tester
Instron 600S Tester
Instron 500S Tester

Figure 2-5
Average Standard Deviation vs. Nominal HRC
Common Process Samples for 4 consecutive lots per HRC Level
Instron 600 Measurements & NIST Deadweight Measurements
25 measurements per Large grade test block

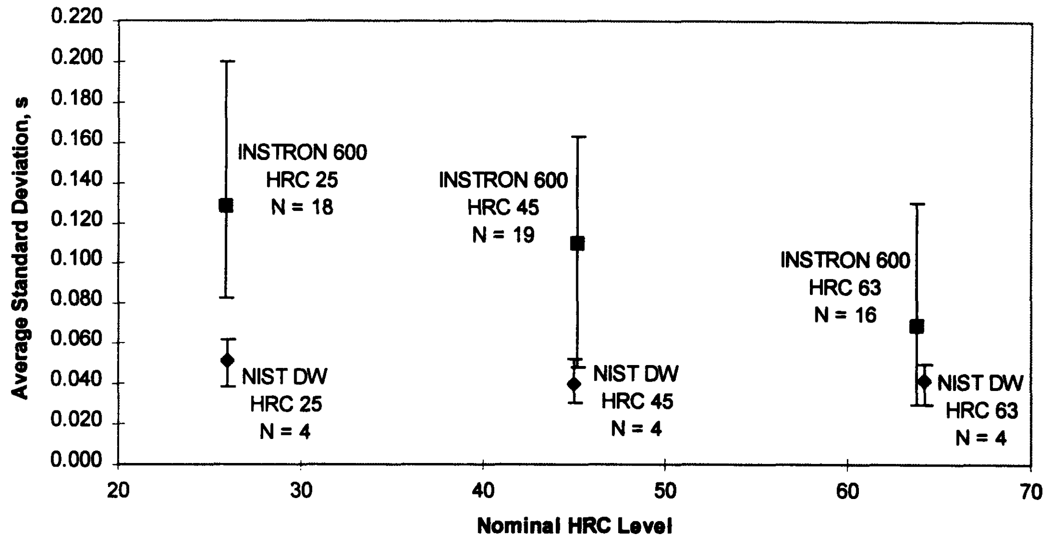
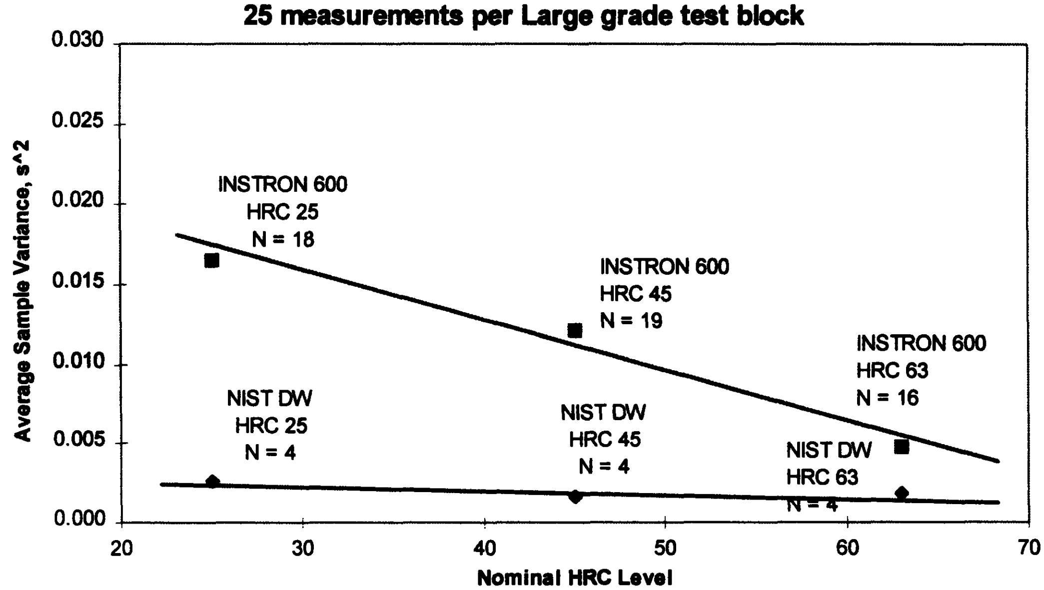


Figure 2-6
Average Sample Variance vs. Nominal HRC
Common Process Samples of 4 Consecutive Process Lots per HRC Level
Instron 600 Measurements & NIST Deadweight Measurements
25 measurements per Large grade test block



2.4 Modeling the Components of Variances

'The variance of the sum is the sum of the variances'. This convenient statistical heuristic holds true only if the two populations being summed are independent. The assumption of independence is currently applied in many gage capability studies, most notably the GRR used for Rockwell hardness testing [3].

For the assumption of independence, the total measurement variation may be expressed in its generic form as a model of sums of variances [18]:

$$\sigma^2_{\text{total measure}} = \sigma^2_{\text{block}} + \sigma^2_{\text{tester}} \qquad (2)$$

In our case the individual quantities, σ_{block} and σ_{tester}, cannot be measured directly. In addition, the assumption of independence is not guaranteed. Therefore, in order to isolate the variation contributions of block and tester embedded in the total measurement variation, each 'quantity must be modeled as a random, and estimated by a prediction interval' [10].

The general form for the components of measurement variation requires a covariance term for the potential correlation between block and tester,

$$\sigma^2_{\text{total measure}} = \sigma^2_{\text{block}} + \sigma^2_{\text{tester}} \ (+\ 2\,\rho\,\sigma_{\text{block}}\,\sigma_{\text{tester}}) \qquad (3)$$

The correlation coefficient, ρ , is thus zero for the assumption of independence. For perfect, positive correlation, ρ equals 1 and for perfect negative correlation ρ is -1. Hence, the substitution of ρ defines estimation intervals for $\sigma^2_{\text{total measure}}$ [10].

Positive correlation, $0 < \rho < 1$: $\quad \sigma^2_{\text{block}} + \sigma^2_{\text{tester}} < \sigma^2_{\text{total measure}} < (\sigma_{\text{block}} + \sigma_{\text{tester}})^2 \qquad (4)$

Negative correlation, $-1 < \rho < 0$: $\quad (\sigma_{\text{block}} - \sigma_{\text{tester}})^2 < \sigma^2_{\text{total measure}} < \sigma^2_{\text{block}} + \sigma^2_{\text{tester}} \qquad (5)$

2.5 Estimates of Variation Components using NIST Deadweight Data

The components of variation models can be used to determine reasonable estimates of the relative block and tester standard deviations for a commercial Instron tester system by the following procedure:

The variances of the NIST deadweight tester measurements may be similarly modeled as:

$$\sigma^2_{\text{DW measure}} = \sigma^2_{\text{block}} + \sigma^2_{\text{DW tester}} \ (+\ 2\,\rho\,\sigma_{\text{block}}\,\sigma_{\text{DW tester}}) \qquad (6)$$

From this model it can be seen that the block variance is at its maximum possible value when the tester variance is zero, such that

$$(\sigma^2_{\text{block}})_{\max} = \sigma^2_{\text{DW measure}} \quad \text{for assumption of } \sigma_{\text{DW tester}} = 0 \qquad (7)$$

or $\quad (\sigma_{\text{block}})_{\max} = \sigma_{\text{DW measure}}$

Note that the maximum condition holds true even if there is a certain degree of dependence e.g. $\rho > 0$.

If the unknown tester is used to measure a *common block* as the NIST deadweight tester, the unknown tester variance can be estimated by a prediction interval. Treating the maximum block standard deviation as known by the standard deviation of the NIST deadweight measurements, the unknown tester variance can be expressed in terms of the measurement variation and block variation using the common block. The nominal, maximum and minimum tester variation therefore is given based on condition of correlation between block and tester:

Nominal, independence, $\rho = 0$: $$\sigma^2_{tester} = \sigma^2_{total\ measure} - (\sigma^2_{block})_{max} \quad (8)$$

Lower Bound, positive correlation, $\rho = +1$: $$\sigma_{tester} = \sigma_{total\ measure} - (\sigma_{block})_{max} \quad (9)$$

Upper Bound, negative correlation, $\rho = -1$: $$\sigma_{tester} = \sigma_{total\ measure} + (\sigma_{block})_{max} \quad (10)$$

Hence, by leveraging the low measurement variation found in the NIST deadweight testers, the variation of the common test block and an Instron tester can be determined within prediction intervals.

2.6 Graphical Analysis for Estimating Tester Variation

The reader will note that the expressions defined for the tester variation are well suited for interpretation of graphs of $\sigma^2_{total\ measure}$ vs. hardness or $\sigma_{total\ measure}$ vs. hardness , in which both NIST deadweight and unknown (600) tester measurements are plotted. Refer to Figures 2-3 through 2-6.

The minimum tester variation is simply the vertical difference between the curve for $s_{total\ measure}$ and the curve for $s_{DW\ measure}$.[4]

The nominal tester variation is simply the vertical difference between the curve for $s^2_{total\ measure}$ and the curve for $s^2_{DW\ measure}$.

2.7 Block and tester variation components determined from sample data

This components of variation model was applied to the measurement data of Appendix C presented earlier in Figs. 2-3 to 2-6 using the 600, 500 and NIST deadweight testers . Sample standard deviations variances were used as unbiased estimators of the model's true standard deviations and variances.

Tables 2.1 and 2.2 present the results from applying the model to the 600 and 500 testers where the same block was measured by NIST (deadweight tester) and Instron. The results shown in Figure 2-5 and 2-6 for cross-referenced samples from the same process run of four consecutive

[4] The sample standard deviation, s, is an unbiased estimator of the true standard deviation, σ.

batches were analyzed and are presented in Table 2.3. The test blocks of this study were solely derived from the Large grade process.

Note that the author purposely abandons calculating the upper bound on tester variation for the condition of perfect negative correlation. This is justified on the basis that the test block standard deviation is conservatively estimated on the assumption that the NIST deadweight tester has zero variation contribution. Only the nominal and minimum tester variations are thus pursued.

2.7.1 Quantifying component contribution to total measurement variation

The relative contribution of tester and block components to total measurement variation can be calculated in two ways depending on the condition of tester and block correlation.

For nominal condition of independence, the relative variation contributions are defined in terms of variances:

$$\sigma^2_{\text{total measure}} = \sigma^2_{\text{tester}} + (\sigma^2_{\text{block}})_{\max} \qquad (11)$$

Hence, the contribution fraction for nominal tester variation is, $\sigma^2_{\text{tester}} / \sigma^2_{\text{total measure}}$ such that

$$\sigma^2_{\text{tester}} / \sigma^2_{\text{total measure}} + \sigma^2_{\text{block}} / \sigma^2_{\text{total measure}} = 1$$

For positive correlation of block and tester, the lower bound condition on tester variation, the relative variation contribution is defined in terms of standard deviations:

$$\sigma_{\text{total measure}} = \sigma_{\text{tester}} + (\sigma_{\text{block}})_{\max} \qquad (12)$$

For this case, the contribution fraction for minimum tester variation is, $\sigma_{\text{tester}} / \sigma_{\text{total measure}}$, such that,

$$\sigma_{\text{tester}} / \sigma_{\text{total measure}} + \sigma_{\text{block}} / \sigma_{\text{total measure}} = 1$$

Most conventions for gaging relative variation contribution to total measurement error use fractions of standard deviations [See 10% rules, 2.9.1 and 2.9.2]. Standard deviation fractions are representative of the condition of negative correlation and minimum tester variation.

The author will remain with convention for purposes of comparison and will reference contribution fractions of variation in terms of *standard deviations* (Eqn. 12).

Table 2-1

Variation Components for 600S Tester using common Large grade block

										Nominal Tester Variation				*Minimum Tester Variation*	
HRC Nominal	Block Serial Number	n_{DW}	STD. DEV. DW measure	VAR. DW measure	MAX. STD. DEV. block	MAX. VAR. block	n_{600}	STD. DEV. 600 measure	VAR. 600 measure	VAR. 600 tester	STD. DEV. 600 tester	$VAR_{tester}/VAR_{measure}$	$s_{tester}/s_{measure}$	STD. DEV. 600 tester	$s_{tester}/s_{measure}$
30	95I30005	68	0.067	0.00449	0.067	0.00449	30	0.161	0.0259	0.0214	0.146	0.83	0.91	0.094	0.58
40	95I40004	76	0.068	0.00462	0.068	0.00462	30	0.091	0.0083	0.0037	0.060	0.44	0.66	0.023	0.25
50	95I50005	75	0.032	0.00102	0.032	0.00102	30	0.060	0.0036	0.0026	0.051	0.72	0.85	0.028	0.47
60	95I60001	75	0.020	0.00040	0.020	0.00040	30	0.076	0.0058	0.0054	0.073	0.93	0.96	0.056	0.74

Table 2-2

Variation Components for 500S Tester using common Large grade block

										Nominal Tester Variation				*Minimum Tester Variation*	
HRC Nominal	Block Serial Number	n_{DW}	STD. DEV. DW measure	VAR. DW measure	MAX. STD. DEV. block	MAX. VAR. block	n_{500}	STD. DEV. 500 measure	VAR. 500 measure	VAR. 500 tester	STD. DEV. 500 tester	$VAR_{tester}/VAR_{measure}$	$s_{tester}/s_{measure}$	STD. DEV. 500 tester	$s_{tester}/s_{measure}$
30	95I30005	68	0.067	0.00449	0.067	0.00449	30	0.170	0.0289	0.0244	0.156	0.84	0.92	0.103	0.61
40	95I40004	76	0.068	0.00462	0.068	0.00462	30	0.141	0.0199	0.0153	0.124	0.77	0.88	0.073	0.52
50	95I50005	75	0.032	0.00102	0.032	0.00102	30	0.120	0.0144	0.0134	0.116	0.93	0.96	0.088	0.73
60	95I60001	75	0.020	0.00040	0.020	0.00040	30	0.127	0.0161	0.0157	0.125	0.98	0.99	0.107	0.84

Table 2-3

Variation Components for 600R Tester using Large grade block samples from a common process (4 consecutive batches)

											Nominal Tester Variation				*Minimum Tester Variation*	
HRC Nominal	N_{DW}	n_{DW}	STD. DEV. DW measure	VAR. DW measure	MAX. STD. DEV. block	MAX. VAR. block	N_{600}	n_{600}	STD. DEV. 600 measure	VAR. 600 measure	VAR. 600 tester	STD. DEV. 600 tester	$VAR_{tester}/VAR_{measure}$	$s_{tester}/s_{measure}$	STD. DEV. 600 tester	$s_{tester}/s_{measure}$
25	4	25	0.052	0.00270	0.052	0.00270	18	25	0.129	0.0166	0.0139	0.118	0.84	0.92	0.077	0.60
45	4	25	0.040	0.00160	0.040	0.00160	19	25	0.110	0.0121	0.0105	0.102	0.87	0.93	0.070	0.64
63	4	25	0.042	0.00176	0.042	0.00176	16	25	0.069	0.0048	0.0030	0.055	0.63	0.79	0.027	0.39

NOTES:

1. DW = NIST Deadweight Tester
2. Nominal Tester Variation for applying Equation (8).
3. Minimum Tester Variation for applying Equation (9).

2.8 Conclusions from variation component modeling and applied results

The components of variances model was successfully applied with help from the NIST deadweight tester; the results are significant. It is concluded that given the current state of block variation of the Large grade blocks, the tester variation accounts for the majority of the total measured variation. In effect, the tester variation masks our ability to measure the variation attributable to the test block.
It can be concluded that on average the tester variation accounts for 50 to 90 percent of the total measurement variation for the Large grade block, depending on hardness level and tester type/technology. Refer to the $s_{tester}/s_{measurement}$ contribution ratios of Tables 2.1, 2.2 and 2.3. These conclusions are drawn based on the modeling assumption that the NIST tester variation is *zero*, such that it measured block variation directly.

The relative tester contribution to total measurement error may be dependent on the type and technology of the tester, as the 500 tester seems to demonstrates a larger variation contribution. The size of the sample data is insufficient to fully yield this inferred conclusion.

The degree of tester contribution to total measurement variation increases with increasing positive correlation of block and tester. This correlation could not be determined within the scope of this study.

It should be noted that these conclusions of variation contribution cannot be extended to Regular grade test blocks. Subsequent data presented in this study (See Table 5.2) shows a clear response by the 600 tester to what appears to be increased variation of the Regular grade test blocks. In effect, the improvement of the Large grade test block has run into the next set of technological control limitations of the measurement system.

2.8.1 Implications of Components of Variation Modeling

An implication of these results is that the future leverage for reducing the measurement variation of the total system for Rockwell hardness measurement lies in the improvement of the tester technology for reduction of variability. Tester technology encompasses the measurement mechanisms, the indenter, as well as the contribution of remaining operator influence (See Figure 2-1).

Moreover it can be concluded that the commercial tester systems pose serious challenges for use as measurement instruments for the purpose of process control in the manufacture of high uniformity, Large grade blocks. The SPC control charts are likely not to be able to discern significant process shifts in the block manufacturing process from the natural or systemic variation arising from the tester at low measurement sample quantities e.g. $n = 5$ (Refer to Chapter 8).

2.9 A common rule-of-thumb for the allowable variation of a measuring device

The variation performance quantified herein for the Instron hardness measurement systems can be compared to an industry 'rule-of-thumb'. This heuristics reflect common expectations stemming from the larger set of industrial measurement devices used to control manufacturing processes.

For purposes of process control of a manufactured product, the tester variation is defined as a fraction of the total measurement variation in testing the product [18]:

$$s_{tester}/s_{total\ measure} \leq 0.10$$

In other words, the standard deviation of the tester system must account for less than 10% of the standard deviation for the readings of process samples (product) being measured. This relative fraction is also applied to GRR studies applied to Rockwell hardness testers [2].

In this case study, our product of interest is the test block developed and manufactured by Instron for use as a calibration reference standard. Since the variation contributions $s_{tester}/s_{total\ measure}$ ratios given in Table 2.1 and 2.3 are generally greater than 0.5, this rule-of-thumb would conclude that the commercial tester system is not suitable for process control of Large grade test blocks.

In most industrial environments this rule of thumb is a lesser issue for Rockwell testing since the manufactured parts tend to have a much higher degree of hardness non-uniformity than a specially-manufactured test block. As a result, the variation contribution of the product substantially increases the total measurement uncertainty.

The relative assessment of the current state of tester variation by this rule-of-thumb indicates that more sophisticated tools and means of analysis are required for addressing the influence of measurement noise in process control of high-quality test blocks. The improved tools may require added rigor, particularly in the application of statistical methods, to yield improved utility.

Chapter 3 Modeling the Propagation of Uncertainty in the Reference Standard

This chapter introduces a conceptual framework for the uncertainty of the reference standard, as it is translated in the test block. This systemic framework models the interactions in the calibration and customer use of the reference standard. From this framework an improved metric is introduced in Chapter 5 that relates the customer's calibration needs to the variation seen in both standardizing block calibration and customer tester validation.

3.1 The standardizing calibration process of the parent system

In the previous chapter a simple model was introduced that aggregated the sources of measurement variation, as the tester, the indenter, the operator and the test block being measured. These sources of variation all come to bear when Instron calibrates a test block for use as a reference standard. The variation can be witnessed by a series of measurements as the measurement values vary in magnitude. The average of the series, $\overline{X}$, typically n = 6 measurements, is recorded on a calibration certificate and engraved on the test block. Refer to Appendix D for a sample certificate. This value is assigned a specification tolerance, δx, that reflects the error in the average $\overline{X}$ due to the sources of variation. The specification half-tolerance, δx, is primarily dictated by ASTM E-18, although the manufacturer may choose a lower tolerance [4]. The goal is to bound the net variation from all of the sources in calibration. A metric for variation, the sample range, is also recorded.

3.2 The tester validation process by the customer/user

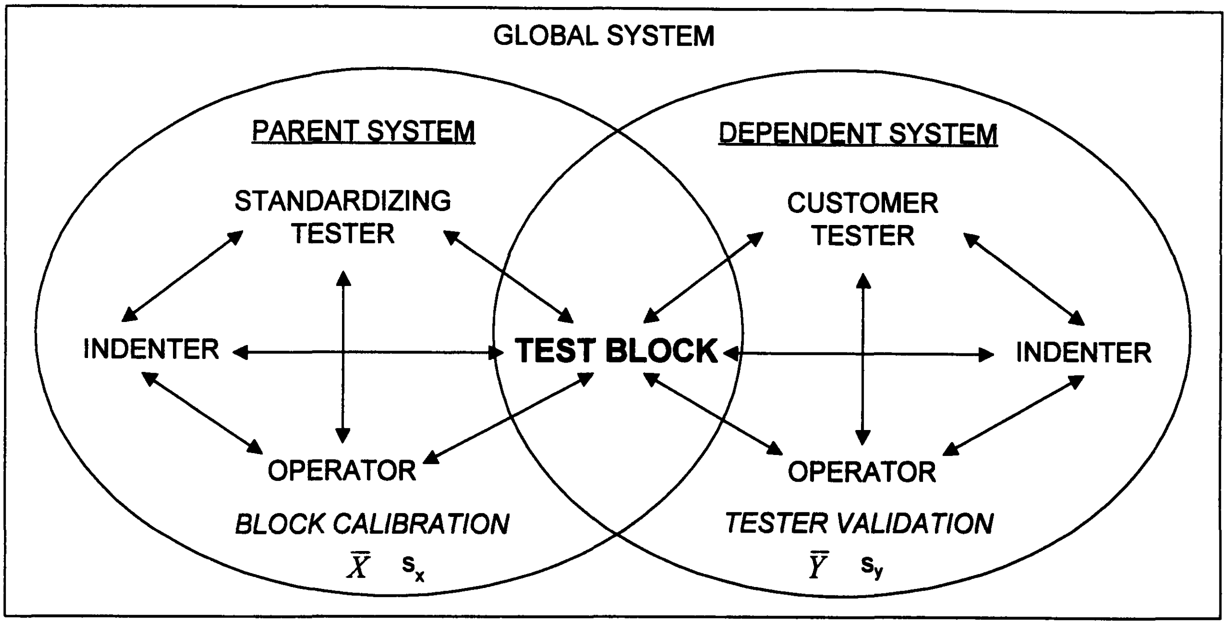
In the process of validating their Rockwell tester, the customer measures the same test block as a reference of the 'true' hardness standard. Again, the same sources of variation this time from a different system all play a role in the variation of hardness readings. The set of customer readings, usually of n=5 or smaller, also yields a measurement average, $\overline{Y}$. Due to the relative sources, both in calibration of the reference standard and measurement of the reference standard, it is unlikely that the two averages $\overline{X}$ and $\overline{Y}$ are exactly identical. In the ideal, $\overline{X} = \overline{Y}$ for perfect validation. However, acceptable validation occurs when $\overline{Y}$ is measured to lie within the tolerance, δx, specified on $\overline{X}$.

3.3 The Global System Model for the Rockwell Hardness Reference Standard

The sources of variation and their interaction of the standards calibration and referencing process are depicted in Figure 3-1. The author terms the standardizing system in block calibration as the *parent system*. The customer validation system is referred to as the *dependent system*. Note that the test block, the reference standard, is common to both parent and dependent systems. Both parent and dependent systems are linked within a *global system* in the capability of producing two measurement averages, $\overline{X}$ and $\overline{Y}$, that are equal.

Note that the hardness measurement response of the dependent system is matched to that of the parent system, as depicted for the generic calibration process of Figure 1-2. As was demonstrated earlier, this matching process must be performed for the different hardness levels being evaluated, since hardness variation changes with nominal hardness level.

Figure 3-1: The Global System for the Rockwell Hardness Reference Standard



3.4 Application of the components of variance models to the global system framework

The degree of variation in either the parent system or the dependent system is measured using the standard deviation, s, (or its square, the variance, s^2) of its measurement readings (as opposed to the range).

When the components of variance model introduced in Chapter 2 is applied to the framework of Figure 3-1, we can see that the variation attributed to the test block is common to both parent and dependent systems. Refer to Figure 3-2 below. Note that the conceptual model assumes the condition of independence of Equation (2), such that all component variances can be added to yield each respective total measurement variance.

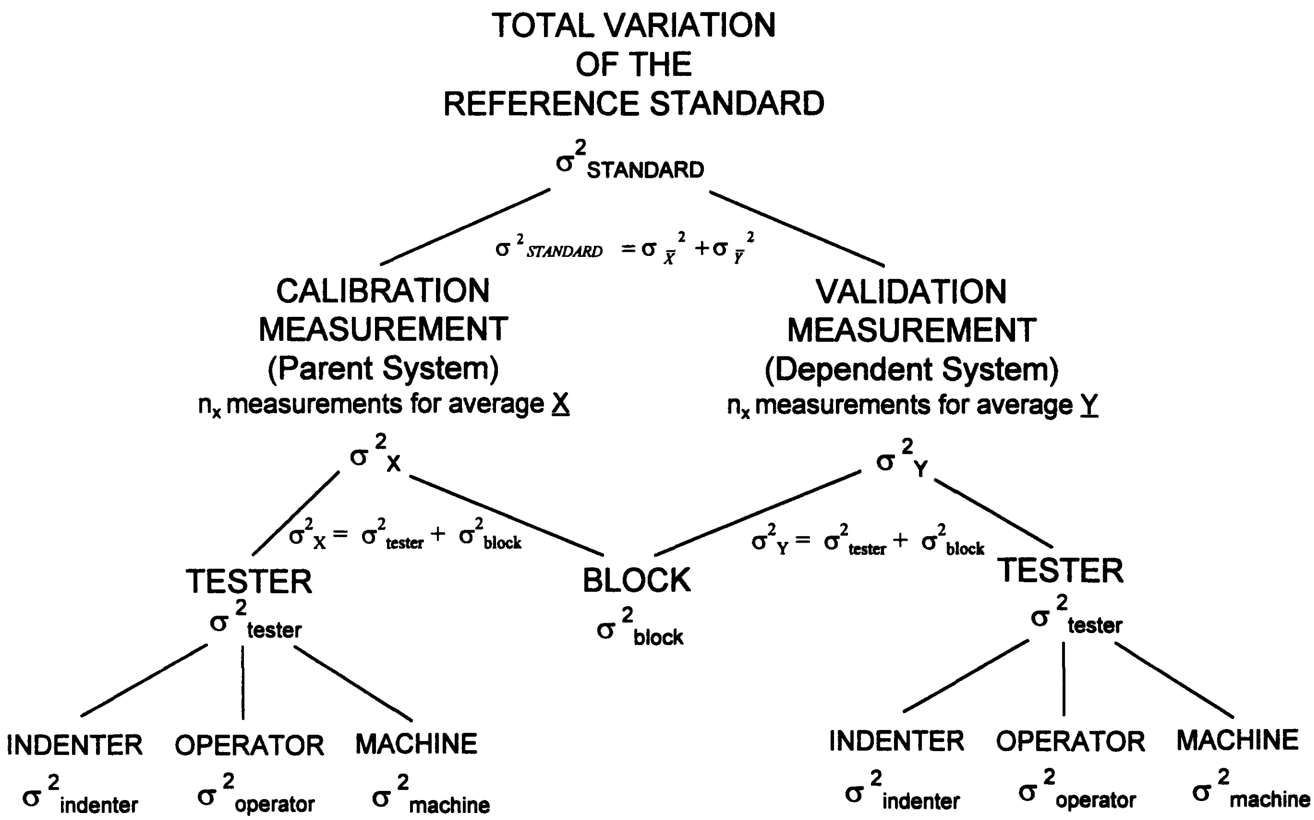
This statistical framework of the sum of variances indicates in general how the ultimate variation of the standard results from the interaction of parent (X) and dependent (Y) systems. Note however that the total variation of the reference standard as perceived by the customer is modeled as the sum of the variances for the *averages* of a set of measurements,

$$\sigma^2_{STANDARD} = \sigma^2_{\bar{X}} + \sigma^2_{\bar{Y}} \tag{13}$$

The *variance of the average* for n independent samples X and Y can be defined by equation (14) [15]:

$$\sigma^2_{\bar{X}} = \sigma^2_{x}/n_x \tag{14}$$

Figure 3-2: The conceptual model for the propagation of uncertainty in the global system as perceived by the test block customer



Therefore, the total combined variation of the reference standard can be written as a combination of the variances of the two measurement averages,

$$\sigma^2_{STANDARD} = \sigma^2_x/n_x + \sigma^2_y/n_y \qquad (15)$$

In practice, in order to obtain sound estimates of the standard deviations, there exists an underlying assumption that the number of measurements is significantly large, n>30. In such a case, the sample standard deviation s can be assumed to equal the true standard deviation, σ.

In addition, the sum of variances of equation (15) assumes that the parent and dependent systems, X and Y, measure *independent* sets of hardness readings despite sampling from the same common block. The underlying assumptions to the statistical models will be addressed in the following Chapter 4.

3.5 Calibration and Validation Objective of the Reference Standard

The calibration and validation objective of the global system depicted in Figures 3-1 and 3-2 is defined by the following statement:

> To reliably determine that the *average* ($\bar{Y}$) of *any* set of measurements performed on the standard test block by the dependent system (Y) *is and will be* within the tolerance band +/- δx of the *average* ($\bar{X}$) of *any* set of readings previously measured by the parent system (X) on the same block.

This objective implies that although two immediate averages of X and Y are measured within the δx tolerance, this is does not suffice to satisfy the global objective of reliably ensuring that *future* measurement averages are also within the tolerance band. In order determine the reliability or confidence regarding future measurements, the inherent degree of variation in both parent and dependent systems must be related to the objective of equating $\bar{X}$ and $\bar{Y}$ within δx. A statistically-based *capability metric*, called the Calibration Capability index, that applies to the global system is introduced in Chapter 5 for that purpose.

The objective also assumes that calibration uncertainty (half-) tolerance, δx , is adequate for the customer vis a vis the precision and accuracy required for their hardness measurements of products in their industrial application. In other words, it is assumed that δx satisfies the customer's calibration needs since the variation of their measured product samples is considerably larger than that defined by δx.

This approach of defining the objective for the hardness reference standard differs from the 10% rule-of-thumb of Section 2.9.2 in that the customer and manufacturer derive an explicit agreement on what degree of uncertainty, δx, is acceptable. This 'agreement' is brokered by the standardizing organizations, such as ASTM, and documented in the form of a standard procedure, e.g. E-18 [4].

The first simplifying step of quantifying the interaction between two independent systems of measurement variation in terms of reliability or confidence is to characterize the nature of the measurement variation in the form of a known and practical probability distribution, such as the Gaussian normal.

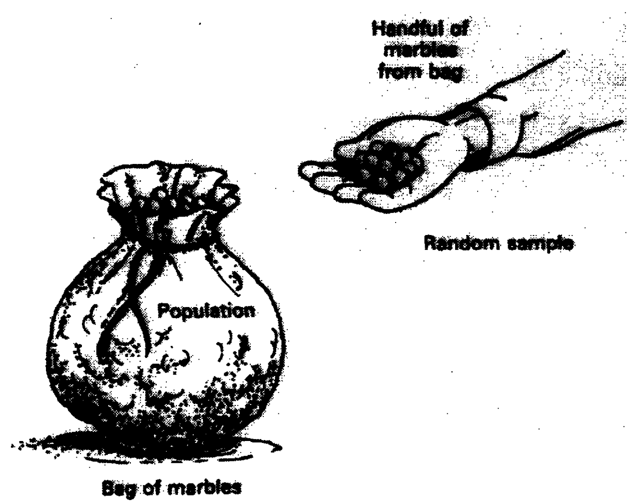
Chapter 4 Discussion of Assumptions for the Statistical Models

The statistical models introduced thusfar are based on key assumptions that must be validated for their application to the environment Rockwell hardness measurement.

The fundamental assumptions are:
- Randomness of individual hardness measurements
- Independence of individual hardness measurements, X
- Independence of hardness measurements from two different systems, X and Y.

The author will draw upon the analogy of a bag of marbles to describe the random and independent sampling from a population as depicted in Figure 4-1 [1]. Each marble is of a different and unique color. Thus, each colored marble represents an individual and unique hardness reading *using a tester machine if infinite precision* that is randomly sampled from the bag population, representing the test block.

Figure 4-1: Bag of Marbles Analogy for a Random Sample taken from a Population [1]



Furthermore, the evaluations of inferential statistics in this study, for example the Calibration Capability index [See Chapter 5] and control chart limits [See Chapter 8], require a characterization of the probability distributions for the individual hardness measurement values. The assumption of a *Gaussian normal distribution* significantly simplifies the statistical arithmetic.

Therefore, this chapter investigates if the hardness measurement values can be approximated by the normal distribution. The nature of the distributions of the measurement constituents of tester and block are explored for purposes of decomposing the sources of variation using the components of variances model of equation (3).

Secondly, the validity of the *Central Limit Theorem* will be evaluated in characterizing the distribution of measurement averages as normal for small measurement subgroup sizes e.g. n=5 or 6. Through proper application of the Central Limit Theorem, we can expect the averages of measurement subgroups to be fairly normally distributed, regardless of the shape of distribution for individual Rockwell measurements as long as the individual (random, independent) measurements are identically distributed.

4.1 Randomness of Test Block Hardness Readings

Randomness in test block hardness sampling is supported by:

- Randomization of measurement location by operator influence
- Random orientation and composition of local material microstructure that determines hardness

4.1.1 Randomization of Measurement Location

ASTM E-18 specifies that indentations are to be 'distributed uniformly' over the surface of the test block in order to establish a basis for random material sampling.

By conducting the Rockwell measurements, the author learned of the difficulty in truly randomizing the measurement location, since there exists a tendency to move away from the point of the prior measurement. However, a well dispersed sequence of indentations is deemed sufficient to provide for the effect of randomization if there should exist some geometrical 'topography' in actual block hardness as shown in Figure 1-4.

The data for the NIST measurements on Large grade blocks followed a prescribed pattern over the full surface of the block. Refer to Appendix E, Figure E-1 for the NIST measurement pattern. This non-random pattern did not significantly influence the superb fit to the normal distribution of the NIST deadweight measurements. Significant patterns in the NIST measurement data were not found in order to invalidate the assumption of random behavior.

4.1.2 Randomness in material micro-structure that determines hardness

The randomness in the particle distributions in the material microstructure under the penetration area of the indenter helps to generate random behavior in hardness measurement. The reader may get an appreciation for the random nature of grain and carbide dispersion within the test block material when inspecting a 1250X enlargement (micrograph) of the measurement surface. Refer to Appendix G for a micrograph that depicts the microstructure of a Large grade, steel test block. The typical average grain size for the specialty steel accounts for an average diameter of less than 10 μm. For a typical indentation diameter of 2 mm, roughly 200 surface material grains influence the resistance to indenter penetration and, hence, the hardness reading.

Thelning [20] describes a empirical model that describes how the microstructural composition determines the (Vickers) hardness of steel based on the relative volume fractions of different iron-carbon phases present[5]. The process of steel-making ensures for a random distribution of individual iron-carbon phases, such as martensite, bainite, retained austenite, that are embodied in individual grains, randomly dispersed amongst other grains of different phases.

The different tones in the micrograph of Appendix G reflect differences in the iron-carbon phases present, as well as carbides, nitrides and small impurity elements in the boundary layers between grains. Note that no two zones appear identical. The distribution of particle sizes and locations of composition tones appears randomly distributed. Hence, it is hypothesized that the indenter interacts with a random set of particles, resisting its penetration, regardless of operator influence. The local hardness response is thus expected to be random.

One way of thinking about the random sampling behavior is using the bag of marbles analogy of Figure 4-1: Each group of material particles subject to the indenter results in a hardness reading. Each hardness reading is analogous to an individual, uniquely colored marble; each marble is randomly sampled from the population of all possible marbles e.g. all groups of particles that make up the test block.

4.2 Independence of Individual Measurements

The condition of independence between individual hardness measurements means that any given hardness measurement X1 does not influence the probability of occurrence of any other hardness measurements/values e.g. X2.

Since the Rockwell test is destructive by nature, every individual hardness measurement samples a different set of microscopic material particles of independent properties, as introduced in Section 4.1.2. Assuming a high degree of measurement precision (1/100th of a Rockwell point), the possible hardness outcomes from a given test block is very large due to the large quantity of possible grouping combinations of material particles for any given indentation.

Using the bag of marbles analogy of Fig. 4-1, if the number of marbles in the bag were small, the removal of a marble *without replacement* would influence the probability of possible outcomes of the remaining marbles[6]. However, if the quantity of marbles in the bag were very large, the removal of any given marble has a *negligible effect* on the probability of other possible readings. This latter situation is the case for large quantity of possible hardness outcomes, such that the test

---

[5] $H_{total} = V_m H_m + V_{p+f} H_{p+f} + V_b H_b$

where, H = Hardness (Vickers)

V = Volume fraction (wt. %)

m = martensite

b = bainite

p+f = pearlite + ferrite

Note that the basis of resistance to penetration is common to both Vickers and Rockwell hardness testing.

[6] For R discrete samples from Q possible outcomes *without replacement*, the probability of occurrence of any remaining sample is 1/(Q-R).

block hardness readings can be modeled as a continuous frequency/probability distribution and the individual measurements may be assumed to be independent.

4.2.1 Effects of Previously Made Indentations

A basic concern is that an interaction between measurements may result due to a local strain-hardening effect from the plastic material deformation of a prior indentation [3]. To avoid this possibility of indent interaction, several standards organizations have published guidelines on the minimum spacing between indentations on the surface of the hardness test block: ASTM E-18 specifies a minimum distance of two and a half indentation diameters from the edge of a prior indentation [4]; ISO specifies a slightly greater spacing of four times the mean diameter of the indentation between the centers of two adjacent indentations [3, SR-1, SR-4].

All test block measurements taken in this study conformed to the ASTM E-18 spacing requirement. It is assumed that this spacing suffices to avoid the effects of previously made indentations. Statistical tests for independence (e.g. Chi-Square test [16]) between sample groups were not applied due to insufficient quantity of sample measurements.

With application of these minimum spacing requirements, block measurements are constrained by a maximum allowable number of indentations that can be placed on the limited surface area of the block without theoretical interaction. Indentations into softer material (e.g. HRC 25) will have larger mean diameters and hence a smaller number of measurements per surface area than that of a harder material (e.g. HRC 63).

The maximum allowable number of measurements is representative of the upper limit of n that is used to define the true block average, μ_x, and its true standard deviation, σ_x, for the test block. A maximum measurement quantity of approximately 120 to 150 indentations can comfortably be placed on a test block of 2.5 in. diameter surface area (4.0 in.^{2}) for hardnesses from HRC 25 to HRC 63 with conformance to the standards rules for interaction avoidance.

4.2.2 Potential Effects of the Tester

Note that independence also implies that the Rockwell tester machine does not exhibit transient behavior within the sample series of measurements. For example, if profuse wear of the tester components was the cause of a time-based drift in hardness readings, the individual measurements could not be assumed to be independent, since they would be correlated by time or sequence.

The author was not able to discern such any transient dependence within a given sample set of measurements, regardless of tester or block type combination.

4.3 Independence of Measurements by Two Systems on the Same Block

The variation of the reference standard of equation (13) is based on the assumption that X and Y are independent. The reader may ask: How can the measurement data sets X and Y be independent if they are sampling the same common block ? Dependence between data sets represents an added complexity for many measurement devices that sample a common feature (e.g. micrometer measuring a part dimension) or that respond to identical material states (e.g. ohmmeter measuring electrical resistor) [16].

However, since the Rockwell test is destructive by nature, no two sets of subsequent measurements can be taken on the same sample of material particles of independent properties as introduced in Section 4.1.2. Just as individual measurements from the same tester are independent [See Section 4.2], if follows that individual measurements from two different tester systems are independent. Because the samples X and Y are taken using two separate tester systems, there is also no possibility for correlation between tester effects. For the high level of measurement precision that we are interested in (e.g. within 1/100th of a Rockwell point), the measurement sets X and Y from two testers are thus assumed independent.

4.4 The Distribution of Combined Measurement Variation

The components of variances model for hardness measurement variation of equation (13) combines the variation sources of *tester* and *block*. The block variance (s^2_{block}) results from the random locational sampling of the actual geometrical distribution of block hardness; s^2_{block} is thus the measure of block non-uniformity. What is the expected form of the distribution of sampled block hardness, described by s^2_{block} ? How does this distribution form interact with the distribution of tester variation ?

4.4.1 The Uniform Distribution for Block Hardness

The frequency distribution for hardness can be determined on the basis of geometric considerations in the probability of hitting a zone K1 of particular hardness X1. Refer to Fig. 4-2 below. Suppose that we have a plane region of area K and a portion of it with area K1, corresponding to hardness X1. Then the probability of hitting the region of area K1 by a random sample is equal to K1/K, assuming that the probability of hitting the plane region K is unity [39]. This geometric basis for the probability of a hardness reading can thus be extended to the hypothetical hardness distribution into numerous hardness zones as depicted in Figure 1-5.

A uniform distribution as depicted in Figure 4-3 is generated if all the possible hardness zones are very small (e.g. smaller than indenter penetration area) such that they can be assumed to be of equal area. For zero measurement error and high precision, each indent hardness reading has approximately equal probability of occurring as any other reading.. The uniform distribution is assumed to be the case for test block hardness samples if:

- indentations from the same indenter are approximately of equal size
- each indent area is comprised of a unique group of microscopic material particles that define its hardness [See Section 4.1.2]
- the material particles are randomly distributed

The author distinguishes between the *random sampling distribution* of the actual block hardness from *the geometrical distribution of a slice* through the test block. There is no reason to expect a geometric hardness distribution of a through-section slice to be normal or perfectly uniform in shape.

If the differences in hardness between zones becomes so small that their hardnesses are assumed equal (e.g. X1=X2=Xn), then the *ideal* uniform distribution of probability 1 (called impulse function) is attained by a pseudo-single hardness zone, whereby area K1 equals K. In this hypothetical case of ideal uniformity, only a single hardness value is measured in any random sample of the block.

Figure 4-2: Geometric Probability for Hardness Distribution across a Test Block

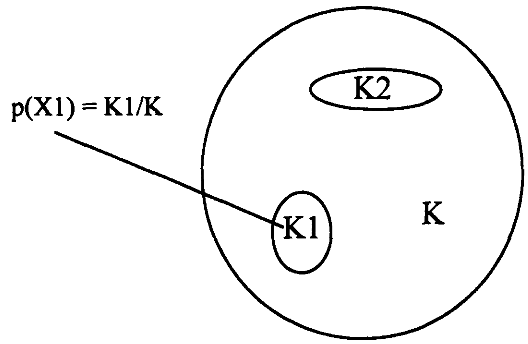
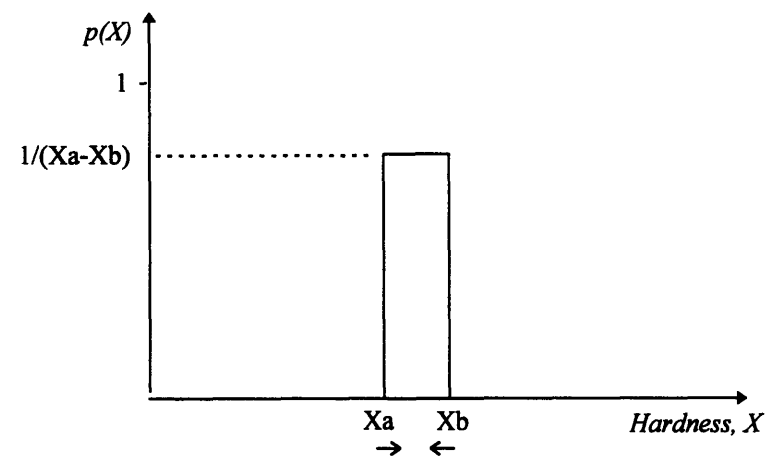


Figure 4-3: Uniform distribution for hardness across the surface of the test block



The expected distribution of material hardness for any series of random samples over the block measurement surface is therefore the *uniform distribution* as depicted in Figure 4-3. It is desired that the range of possible hardness values (Xa -Xb) is a very narrow band[7]. The ideal goal of uniformity is one where Xb is approximately equal to Xa and the frequency of occurrence approaches 1.

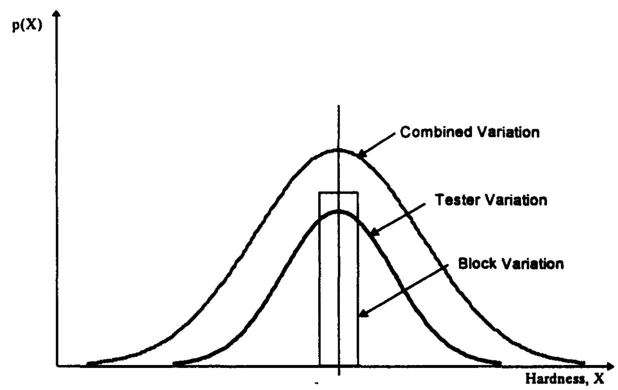
[7] For uniform distribution, the variance, $\sigma^2 = (Xb-Xa)^2/12$ [39]

4.4.2 Combined Block and Tester Distributions

With the effects of *tester* variation on hardness measurements, the actual distribution of hardness by random sampling cannot be directly determined. The best we can currently do is to study the measurement distribution resulting from the lowest measurement variation of the deadweight tester.

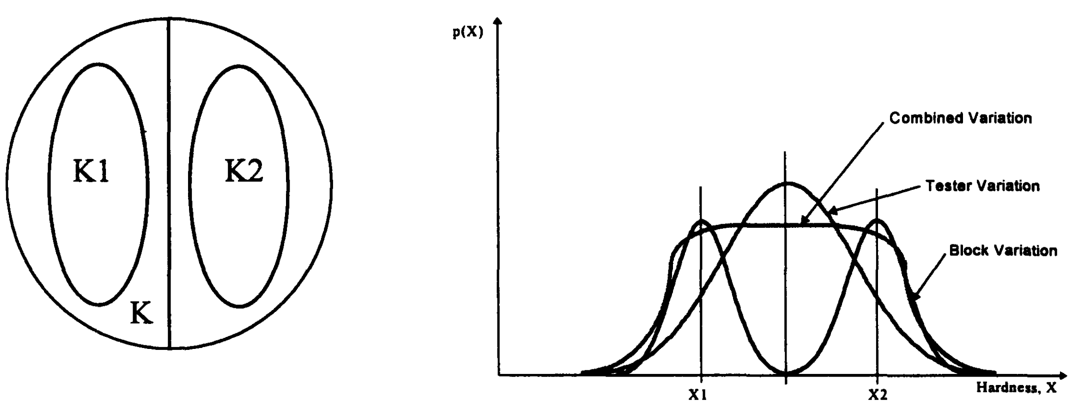
Measurement noise (tester variation) is often modeled as a normal distribution [1] due to the large quantity of simultaneous effects. If the block distribution is uniform and of small hardness range relative to the tester variation, the normal form of the measurement noise will dominate the resulting combined measurement distribution. Refer to Figure 4-4 below.

Figure 4-4: Normal tester variation dominating a narrow uniform block distribution



Now hypothetically suppose that the hardness distribution of the test block was largely bimodal with two distinct and dominant hardness outcomes, X1 and X2 as shown in Figure 4-5. In such a case, the distribution of hardness would be bimodal in shape. The combined measurement variation would thus follow some combination of the bimodal and normal distributions.

Figure 4-5 Bimodal Block Hardness Distribution combined with Normal Tester Variation



Thus, if analysis of the resulting hardness measurements displays a normal distribution, we may conclude that we are largely witnessing several possible effects:

- the normal tester variation is superimposed on a normal block distribution. Combined normal distributions yield a normal distribution.
- the normal tester variation dominates a narrow block distribution that is approximately uniform ($\sigma_{tester} > \sigma_{block}$).
- there are no gross variations in block uniformity to cause the combined distribution to be non-normal e.g. the different zones of Fig. 4-2 are small and numerous and the differences in their hardnesses are small.

4.5 Literature Study on the Random Normality of Test Block Measurements

In a 1984 literature study on hardness test blocks and indenters, the OIML summarized published results addressing the question of normality for the distribution of individual hardness measurements [3]. Their synopsis is to some extent repeated here.

Many of the studies explored if there exists a geometrical 'topography' of hardness over the surface of the test block. The practice of how the individual measurements are randomized by the operator then becomes a central issue.

4.5.1 The Normality Cons

Yamamoto [3, Y-4] and Cutka [3, C-1] found cases where there exist fields of concentric contours of common hardness numbers on the test block, in some cases parallel contours. It is cited that only in rare occasions do the hardness readings not follow a discernible pattern and can be considered randomly distributed over the surface. This conclusion supports the representation model of Figure 1-4 for a geometrical hardness dispersion.

Yamamoto [3, Y-4] further concluded that only in rare cases can the distribution of hardness on the surface of a test block be assumed normal, even when test points are selected at random; thus the mean of the population cannot be estimated accurately.

Marriner [3, M-4, M-7] shares the opinion that distribution of hardness values over the test block is not Gaussian. However, 'the sampling variability from the test block contributes largely to the spread of measure values', as determined for subgroups of n=10 randomly disposed over the whole surface. Note that Marriner distinguishes between normal measurement behavior, resulting from normal measurement noise (tester variation) and actual block hardness sampling distribution.

4.5.2 The Normality Pros

Cutka [3, C-1] created a histogram that supported the hypothesis of a normal distribution of values with certain limiting conditions, such as distributing the applied indentations across the surface of the block.

Petik [3, P-11] performed statistical tests on 212 sets of n=25 measurement values. The hypothesis of normal distribution of the measurement values over the 212 data sets could not be rejected at the 5% significance level.

4.5.3 Partial conclusions from the literature

The OIML study [3] concludes on the basis of disagreements in these references regarding normality that the question of normality for 'hardness values' is still disputed and thus inconclusive. "Consequently some restrictions are to be employed at applying certain statistical methods for the evaluation of hardness standardizing measurements"[3]. Clearly, the dispersion of indentations across the block surface is taken as a necessary requirement for achieving randomization of measurement values.

In addition, there appears to be a conflict of trying to characterize the expected sampling behavior due to the block's inherent geometrical hardness distribution versus the response behavior of the measurement system under the influence of measurement noise (tester variation). The effect of measurement noise is that it masks the true sampling distribution of block hardness. Hence, while the measurement variation may behave normal, it does not preclude that the block's hardness distribution from random sampling is normal.

4.6 Statistical Methods in Literature Reflect Implicit Assumptions of Normality

The statistical methods that other standardizing organizations and industry experts apply on hardness measurement values also reflect their implicit assumptions of the probability distributions, even though their assumptions may or may not be explicitly stated.

NAMAS, an independent British accreditation organization, in their document NIS 0406 'Hardness Testing Equipment' [9] define a methodology for the expression of uncertainty for the measurement average. In addition, a calibration uncertainty between the mean value from the test block certificate and the sample average is introduced with formulae. The methodology applies the tabulated t-statistic at a 95% confidence level to define the uncertainty for the measurement average. Only examples of measurement quantities less than 10 are given. Hence, NAMAS implicitly assumes that we are sampling from a normal distribution and applying the t-statistic to correct for small sample quantities.

A 1993 Cornell Master's project with Wilson Instruments Delcourt et al define the components of variability on the basis of :"the readings in order to label a block are assumed to be from a normal distribution with mean μ and variance σ^2_L" [12].

Finally, the GRR studies applied to Rockwell hardness testing apply conversion factors on the sample Range to provide estimates of the standard deviation; these factors are implicitly based on the normal distribution for a defined sample size [2].

4.7 Goodness-of-Fit of the Normal Distribution for Test Block Hardness Measurements

The author studied the hypothesis of the Gaussian normal distribution for hardness measurements in sampling a test block of the Rockwell 'C'-scale using several analysis tools:

- Q-Q plots of the individual measurements converted into standard normal variables (Qi) vs. standard normal variables (Zi) [14]. For normality the plotted points lie on a 45 degree slope through the origin.
- Measures of *Skewness* (γ_1) and *Kurtosis* (γ_2) describe the shape of the underlying distribution. Bell-shaped frequency distributions will have a coefficient of skewness $\gamma_1 = 0$ for symmetry and a coefficient of kurtosis $\gamma_2 = 0$ for the relative concentration along the tails or 'degree of peakedness' [16, 39].[8]
- Histograms in order to match the shape the bell- curve of a normal distribution. This method is invoked when the round-off error of the tester read-out inhibits the interpretation of the Q-Q plot, e.g. measurements by the 500 tester with one significant digit (X.X).
- Comparison of the range to standard deviation (R/s) behavior of the measurement data against the theoretical behavior for a normal distribution.

Note that the applied tools do not include the chi-squared test for normality using parsed data sets [14]. In general, the data sets were not sufficiently large (less than 75 points) to draw statistically significant conclusions.

In evaluating the Q-Q plot, normality is confirmed if the data points lie on a 45 deg. slope through zero. In addition, the 'S'-shape of the sloping curve gives some indication if the variance is over- or underestimated [14]. If the data point line deviates from the zero, it is an indication that the mean is in error. For a normal distribution the points are tight near zero and their spacing grows the further away from zero.

Devor [14] also cites that the Q-Q plot facilitates relative assessments of goodness-of-fit between several sets of sample data. Hence, the Q-Q plots can be directly compared. Q-Q plots are generally more versatile in their application than histograms as they do not require determination of a rational bin size; both tools, however, require large data samples for significant conclusions.

The hypothesis of the normal probability distribution is not directed at the nature of the test block per se; instead, the distribution of the individual hardness measurements are treated as a response from the entire measurement system, consisting of the tester mechanism, the indenter, the operator influence and the test block. Only the total system response is available to the investigator. Therefore, the measurement distribution is representative of the particular tester system.

The author studied several types of tester systems and two block types over 3 to 4 hardness levels in addressing the normality hypothesis. Three different tester systems were applied to a set of common Large grade blocks at 30, 40, 50 and 60 HRC. In addition, the method of measurement may in certain cases have followed prescribed patterns over the surface of the block. Refer to Figure E-1 of Appendix E; in other cases the measurement locations were randomized.

---

[8] Coefficients of skewness and kurtosis are defined by nth order moments moments (μ_n). See [16] where,

$$\gamma_1 = (\mu_3^2/\mu_2^3)^{0.5}$$
$$\gamma_2 = \mu_4/\mu_2^2 - 3$$

The actual measurement data is provided in Appendix C. Refer to Table 4.1 for a summary of the investigations.

Table 4.2: Summary Table for Normality Goodness-of-Fit Investigations

Nominal HRC	Block Type	Block Serial No.	Tester System (1)	Indent Qty. n	Pattern	Skewness γ_1	Kurtosis γ_2	Q-Q Plot (Fig.)	Histogram (Fig.)
30	Large	95I30005	NIST DW	68	Fig. E-1	-0.01	-0.34	F-1	--
			INSTRON 600S	30	Random	0.44	-0.49	F-2	--
			INSTRON 500S	30	Random	0.29	-0.07	F-3	--
40	Large	95I40004	NIST DW	76	Fig. E-1	-0.40	-0.03	F-4	--
			INSTRON 600S	30	Random	-0.31	-0.60	F-5	--
			INSTRON 500S	30	Random	0.51	-0.10	F-6	--
50	Large	95I50005	NIST DW	75	Fig. E-1	.065	0.30	F-7	--
			INSTRON 600S	30	Random	-0.48	0.00	F-8	--
			INSTRON 500S	30	Random	-0.25	0.12	F-9	--
60	Large	95I60001	NIST DW	75	Fig. E-1	0.15	0.02	F-10	--
			INSTRON 600S	30	Random	0.40	0.57	F-11	--
			INSTRON 500S	30	Random	0.63	3.1	F-12	--
25	Large	I25016	INSTRON 600R	60	Random	0.17	-0.28	F-13	--
45	Large	I45005	INSTRON 600R	60	Random	-0.36	0.85	F-14	--
45	Large	I45006	INSTRON 600R	60	Random	-0.56	-0.15	F-15	--
63	Large	I63020	INSTRON 600R	60	Random	-0.23	1.96	F-16	--
25	Regular	H00128	INSTRON 600S	30	Random	-0.16	1.38	F-17	--
			INSTRON 500S	30	Random	0.27	-0.21	F-18	F-23
45	Regular	G00390	INSTRON 600S	30	Random	-1.11	0.17	F-19	--
			INSTRON 500S	30	Random	-0.53	-0.58	F-20	F-24
63	Regular	R02539	INSTRON 600S	30	Random	-0.52	1.82	F-21	--
			INSTRON 500S	30	Random	-0.73	0.98	F-22	F-25

NOTES:
Tester Types:
NIST DW = NIST Deadweight tester
Instron 600R = Research 600 tester (Model 653 C, Ser. No. 97331502); indenter 95621105
Instron 600S = Standards Lab tester (Model A653RMT-4, Ser. No. 97328502); indenter 95621105
Instron 500 = Standards Lab tester (Model B523R, Ser. No. 80195408); indenter 940191

4.7.1 Results on Goodness of Fit to Normal Distribution

The following findings are drawn from the goodness-of-fit results given in Table 4.1:

- The Q-Q plots of the NIST deadweight tester data for Large grade blocks allows the conclusion that measured NIST deadweight variation is normal for hardnesses HRC 30 to 60 despite the non-random measurement patterns. Due to the low degree of total variation seen earlier for this data in Chapter 2, this provides the closest indication that the variation of the block may itself be approximately normal, although not conclusively [Refer to Section 4.4.2].
- The increased measurement quantity (60 to 76 data points vs. 30) appears to result in an improved fit to the normal distribution. This is evidenced for the Large blocks tested with the 600 unit e.g. serial no. I25016. This improved fit with increased n is expected using the Q-Q plot methodology. [14,15].
- All of the Q-Q plots demonstrate several outliers at the tails that deviate from the general normal behavior. Their cause is unknown. These must be accounted for in use of the measurement data for inferential statistics and in setting up control charts limits.
- The 500 tester results are subject to round-off error in the tester read-out or measurement algorithm to the nearest tenth e.g. XX.X. As a result, 'lumps' of common hardness measurements are seen on the Q-Q plots. At low n=30 data points, this round-off is deemed to significantly impact the quality of the goodness-of-fit tests for the 500 tester. The 500 tester results for the Regular blocks were therefore confirmed with histograms binned at the round-off accuracy level of XX.X for the read-out. These histograms show a general tendency toward normality.
- The Q-Q plots for the 600 testers at higher hardnesses of HRC 60 and 63 also show a slight lumping of 3 to 5 data points. This lumping is also attributed to a rounding or binning in the tester algorithm/read-out due to the lower degree of variation at the higher hardness.
- One pair of data sets for Regular block of serial no. G00390 at HRC 45 shows particularly poor fit results for both 600 and 500 tester at n=30. The 600 tester has both mean and variance errors with poor distribution (skewness). The same poor fit for two different testers may indicate that gross variations in hardness of the Regular block is influencing the normality of the total measurement variation.
- There is no detectable pattern in skewness, either positive (shift left) or negative (shift right).
- In general, the skewness measures indicate that the measurement distributions are fairly symmetrical with the exception of the Regular block, serial no. G00390 measured by the 600S system.
- The goodness-of-fit is not consistent across different testers applied to the same block. There is insufficient data to suggest that the goodness-of-fit is attributable to the true hardness distribution of the block vs. the influence of tester variation.

4.8 R/s Behavior for Sets of Individual Measurements

The range-to-standard deviation ratios, R/s, were calculated for 158 different test blocks that were measured on the common Instron 600R tester set-up. These blocks were of all types, both Regular and Large grade, and constituted a spectrum of hardnesses from HRC 25 to HRC 63. Up to 25 measurements were taken per individual block following a prescribed geometrical pattern. Refer to a Appendix E, Figure E-2.

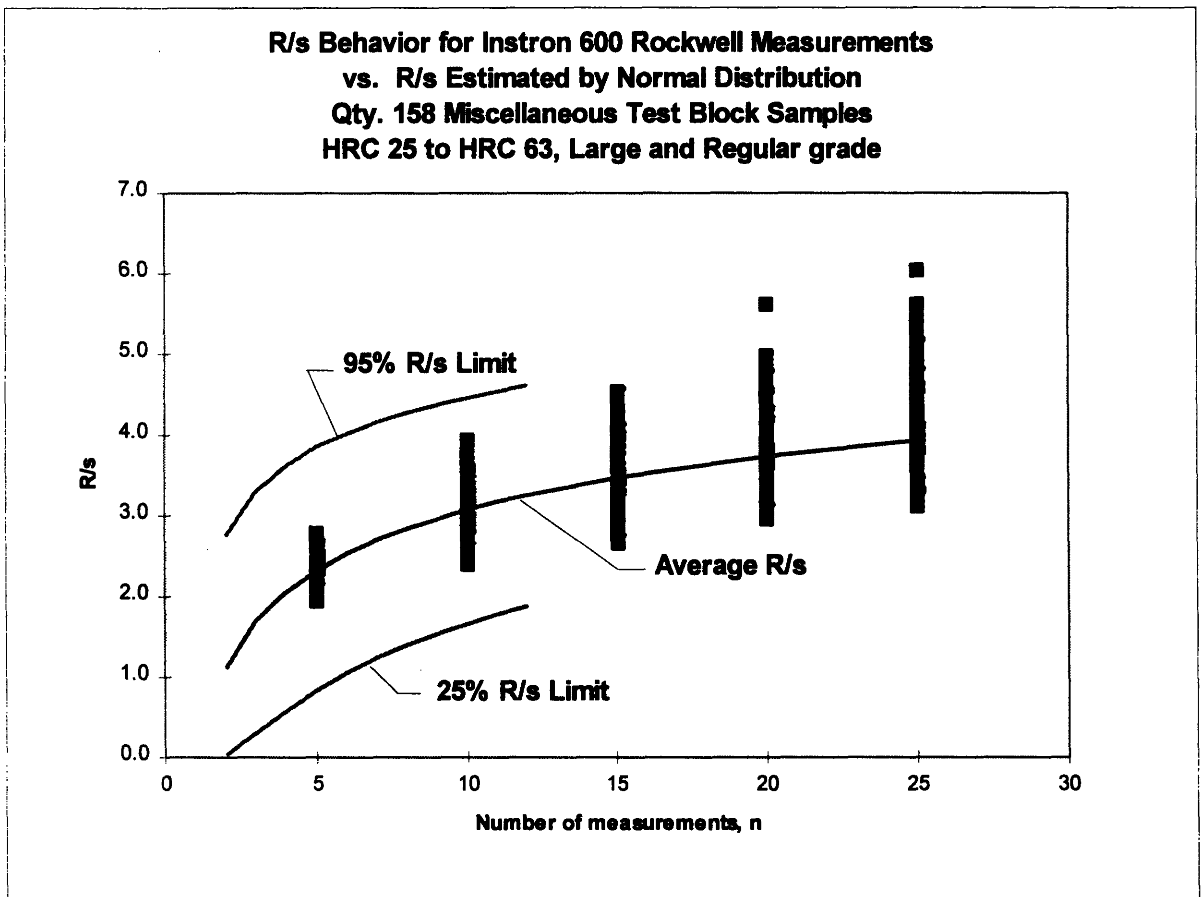
The R/s measurement results are plotted against the predicted R/s values from a normal distribution with respect to number of measurement samples, n [16]. Refer to Figure 4-6 below.

The estimates from the normal distribution are based on confidence limits. For instance, the 95% limit means that 95% of the data points taken from a normal population are expected to lie below the 95% limit curve. It should be noted that the *average* 50% curve for a normal population is the expected R/s value. It also represents the common d2 factor used in calculating the Shewhart control chart limits for range data on the R-chart [16].

The reader will note that the block measurement values are well distributed above and below the 50% curve and generally remain within the 95% R/s limit and lie above the 25% limit. At increased $n \geq 20$ the R/s spread shifts up slightly, such that at n=25 there are 5 blocks that exceeded the 95% limit.

Hence, the R/s behavior of measurement subgroups is representative of that for the Gaussian normal distribution with some indication of non-conforming outliers.

Figure 4-6: R/s behavior of test block measurements vs. the normal distribution



4.9 Conclusions on the Probability Distributions of Individual Measurements

The measurement distribution can be assumed approximately normal

The author concludes that the assumption of normality for the random sampling distribution of individual measurements across the surface of the block is reasonable *on average*. There exists a likelihood of extreme non-normal outliers within each measurement data set, as well as isolated instances of particularly poor goodness-of-fit between measurement trials/data sets.

The block hardness distribution is unknown

A theoretical justification for a uniform distribution for block hardness for random locational sampling can be made [See Section 4.4.1]. The normality of NIST deadweight measurements for the Large blocks fit the model of normal and low measurement noise (tester variation) superimposed on a narrow uniform block hardness distribution. However, their is no empirical evidence to suggest that the Large block hardness distribution is not normal for random sampling. Literature references are not conclusive on the issue.

Non-normal outliers in measurement data sets

The outliers that deviate from the normal distribution at the tails must be accounted for when considering statistical use of test block data of *individual measurements,* such as the Range Capability index of Chapter 7.

Hardness non-uniformity of Regular blocks is unknown

The Regular blocks have not been tested by the NIST deadweight tester of low tester variation. Thus, conclusions on the form of the sampling distribution for actual hardness of the Regular blocks cannot be made. In general, gross variations in hardness across the block will invalidate the assumption of normality. The single example of non-conforming behavior of HRC 45 block G00390 would support the notion of gross hardness non-uniformity.

However, in general the measurement variation for Regular blocks with the influence of measurement noise from commercial Instron testers behaves approximately normal. Therefore, measurement data from Regular blocks can be used for statistical methods and SPC control charting.

Randomization in locational sampling is good practice

In general, the requirement for randomization in measurement location for Large grade blocks is not deemed significant as demonstrated by the patterned NIST data. The theoretical model for a random dispersion of microstructural constituents that define local hardness of Section 4.1.2 supports the notion that randomization is predetermined.

However, given the literature findings for zones of common hardness, the current policy of randomly distributing the measurements across the block surface is concluded to be sound and of good practice. In addition, since the degree of hardness non-uniformity is not known for Regular blocks a well-distributed measurement sequence is important.

R/s measurement behavior behaves markedly normal for small n
For small measurement quantities of less than 10, the R/s behavior, as depicted in Figure 4-6, supports the conclusion that total measurement variation can be practically assumed normal. The implication of such predictable behavior for individual measurements of small sample groups is that estimate conversions between Range R and Sample Standard Deviation s can be made.

4.10 Normality of Averages: The Central Limit Theorem

The Central Limit Theorem states that for n independent observations of X with theoretical true mean and variance of μ_x and σ_x^2, the sum of the averages $\bar{X}$ will be normally distributed with a mean of μ_x and a variance $\sigma_{\bar{X}}^2 = \sigma_x^2/n$, regardless of the distribution of the parent population [15].

Hence, by taking measurement subgroups from the population of all possible block measurements, their averages may be assumed to be normally distributed even when their parent population is not as normal as we would like. Hogg and Ledolter [15] cite that the theorem requires the sample size to be sufficiently large ($n \geq 30$) to approximate the distribution of the sample average X as a normal distribution. In a large-sample setting, sample standard deviations s_x and s_y can be used as sound estimates of true standard deviations σ_x and σ_y. For small subgroups (e.g. n= 5 or 6), the estimates of σ_x and σ_y by s_x and s_y may therefore be in error.

Since most hardness measurement subgroups are less than n=10, the robustness of the Central Limit Theorem may be questioned for most practical applications.

However, Hogg and Ledolter demonstrate through simple simulation sets of random samples from a uniform parent distribution that the C.L.T. results in a distribution of averages that is closely normal for n=5 subgroup sizes [15].

Duncan applies the normal z-test for samples sizes less than 6. He also states that the assumption of normality is fairly robust, because even if universes are moderately non-normal, means and differences of means tend to be normally distributed [16].

The Central Limit Theorem is used for achieving normality of means in Shewhart control charting methods even when samples are small. A sample size of 5 is typical for Shewhart control charts. If the parent populations are non-normal, the averages of small sample sub-groups are assumed normal. As a result, the Gaussian z-statistic is applied for confidence intervals on these subgroups.

Devor et al cite on the basis of Shewhart control charts that the sample size generally should be 4 or more to ensure good approximation to normality. Note that a small subgroup size is desirable for control charts to ensure responsiveness to common-cause variation and to be 'economically appealing' from a collection and measurement standpoint [14].

4.10.1 A Simulation of the Central Limit Theorem on Sample Data

A simple simulation was conducted on the measurement data for the Large block I45005 (See Appendix C). The data was parsed into 12 subgroups of size n=5. A second Q-Q plot was constructed for the subgroup averages. The new Q-Q plot for subgroup averages is compared to the Q-Q plot of individual measurements. The averages Q-Q plot shows a slight improvement in fit compared to the plot for individual measurements. However, the fit for the conclusion of normality is not expected to be perfect, partially due to low quantity of data points.
Note that the actual standard deviation of 0.090 for the distribution of averages over-estimates the theoretical standard deviation of 0.070 of equation (14). Clearly, accurate estimates of the true standard deviation remains an issue for small subgroups.

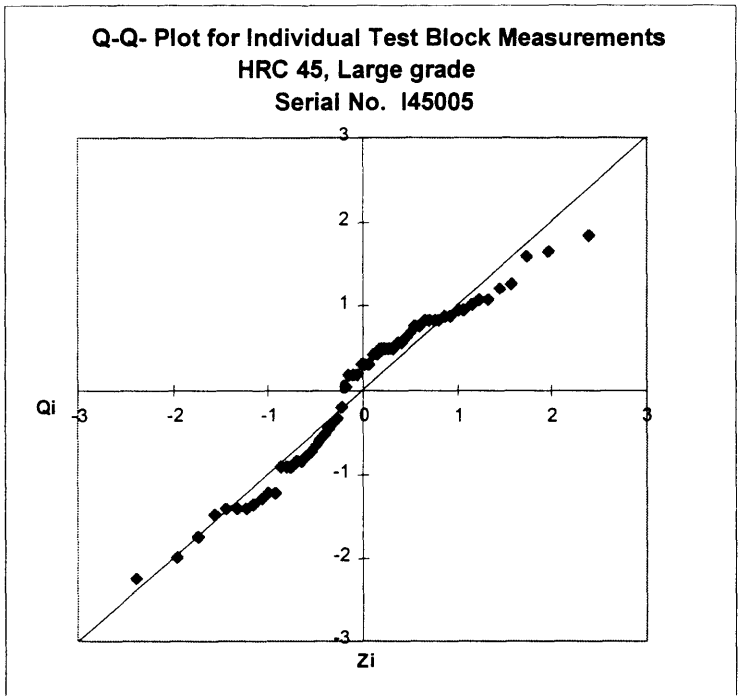
4.10.2 Conclusions for Normality of Subgroup Averages

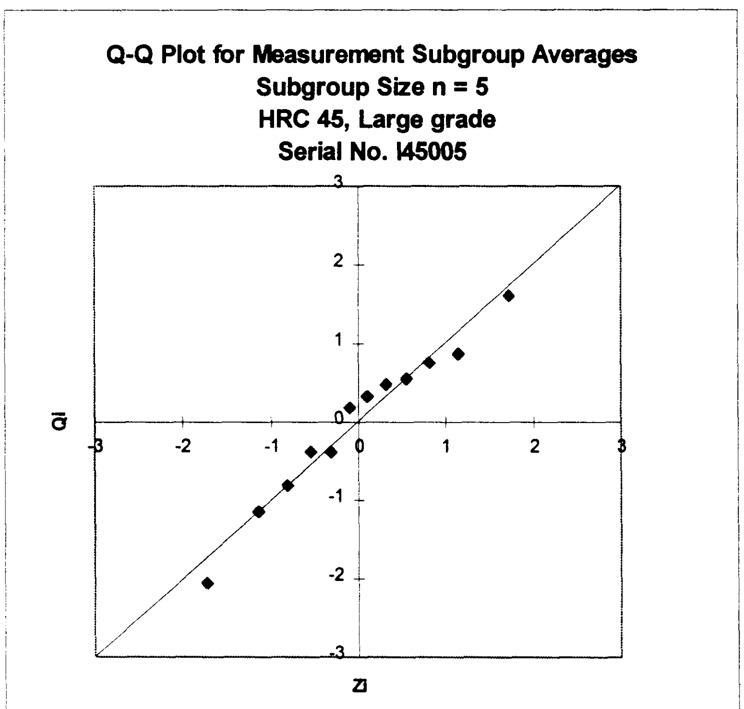
It was demonstrated earlier that individual block measurements tend to behave fairly normal with instances of non-conformance or inferior fit. Thus, if measurement distributions X and Y are moderately normal, then the additional influence of the Central Limit Theorem (even for small samples less than 10) would support the conclusion that subgroup averages, $\overline{X}$ and $\overline{Y}$, and their differences, $\overline{X} - \overline{Y}$, may be treated as normally distributed.

The simulation example of Figures 4-2 did not provide overriding evidence for the effect of the Central Limit Theorem in achieving normality of subgroup averages; this is perhaps a sign of weak independence for this particular data set. Another reason is the limited number of data points available for the evaluation.

Generally the use of subgroup averages in statistical evaluations allows the assumption of the normal distribution due to the Central Limit Theorem. Therefore, the questionable normality of Regular block hardness non-uniformity for low tester variation is not a limiting factor in using normality-based statistical methods that employ subgroup averages. Thus, the capability indices Cc, Cpk and C_R [See Chapters 5, 6 and 7] and conventional SPC control chart limits [See Chapter 8] are founded on a reasonable assumption of normality.

Figures 4-2: Normality goodness-of-fit: Individual measurements (Before) and Subgroup Averages (After)





Part II Capability Metrics for Product and Process Performance

Chapter 5 Development and Definition of the Calibration Capability

The author developed the Calibration Capability index from statistical foundat-ions of hypothesis testing, as is the case with the foundations of the more common process capability indices, Cp and Cpk. This metric gages how well the stated objective of Section 3.5 is fulfilled from the viewpoint of the customer/user.

A hypothesis test is applied for the difference of two population means μ_x and μ_y. These populations reflect measurement data from two separate tester systems, the parent system X and the dependent system Y. Although these populations are derived from measuring the same common block, each tester system yields indentation measurements stemming from two discrete sets of material 'grains', randomly distributed over the block's surface.

The hypothesis test depends on a set of underlying assumptions that were successfully validated in the previous chapter. These assumptions are summarized:

- Independence of the measurement sample populations of the parent and dependent test systems, X and Y.
- Hardness measurements behave as random variables that tend to the normal distribution with increasing sample quantity. Averages of measurement subgroups are expected to conform to the normal distribution due to the Central Limit Theorem.

The following development shows different test and index formulations based on the varying sets of assumptions regarding the estimates of the standard deviations, σ_x and σ_y. The author studies the effect of these assumptions on the basis of two simulation scenarios. The goal is to formulate the Calibration Capability index in its simplest, most practical form without significant penalty for taking liberty in simplifying the assumptions.

5.1 The One-Tailed Hypothesis Test

For condition where $\mu_x > \mu_y$, the null hypothesis to be tested is stated as,

$$\text{H0:} \quad \mu_x - \mu_y \geq \delta x \qquad \text{or} \quad \mu_y \leq \mu_x - \delta x \tag{16}$$

The alternate hypothesis, representing a favorable validation condition, is therefore,

$$\text{H1:} \quad \mu_x - \mu_y < \delta x \qquad \text{or} \quad \mu_y > \mu_x - \delta x \tag{17}$$

The critical rejection region for H0 is defined by:

$$T \leq -t(\alpha;\upsilon) \tag{18}$$

Otherwise, the alternate hypothesis H1 is accepted.

Note that this test is equivalent to testing if the true average μ_y lies within a $-\delta x$ tolerance on the true average μ_x which has a +/- δx tolerance.

This test represents a one-tailed test using the Student's t-distribution, where the test statistic is given by:

$$T = \frac{\overline{X} - \overline{Y} - \delta x}{s_{\overline{X}-\overline{Y}}} \quad \text{where } \overline{X} - \overline{Y} \text{ is positive for } \overline{X} > \overline{Y} \tag{19}$$

List of Variables:

μ_x : theoretical true mean of the parent system measurement population, X

μ_y : theoretical true mean of the dependent system measurement population, Y

t : critical value of the Student's t-distribution

α : significance level or the probability of Type I error[9]

υ : degrees of freedom of the Student's t-distribution

$\overline{X}$: Computed average hardness of n_x measurements of the parent system X calibrating the test block.

$\overline{Y}$: Computed average hardness of n_y measurements by the dependent system Y of the test block in validation use.

δx : Specification half-tolerance on the block calibration average $\overline{X}$ of parent tester system X, as engraved on block by the manufacturer (e.g. $\overline{X}$ +/- δx)

$s_{\overline{X}-\overline{Y}}$: Equivalent standard deviation for the distribution of $\overline{X}$ - $\overline{Y}$

Thus, if

$$\frac{\overline{X} - \overline{Y} - \delta x}{s_{\overline{X}-\overline{Y}}} \leq - t(\alpha;\upsilon) \tag{20}$$

then the null hypothesis H0 (the difference between the measurement averages is *greater* than the half-tolerance) is rejected AND the alternate hypothesis (the difference is less than the half-tolerance) is accepted.

The reader will note that the favorable condition of tester Y validation (within the -δx half-tolerance on $\overline{X}$) is represented by the alternate hypothesis H1.

5.1.1 Absolute value on $\overline{X}$ - $\overline{Y}$

Recalling that the half-tolerance is on both sides of $\overline{X}$ e.g. +/- δx, the condition is evaluated for $\mu_y > \mu_x$ where the +δx half-tolerance is applied to $\overline{X}$. In this case, the hypothesis to be tested is stated as follows:

$$\text{H0:} \quad \mu_y - \mu_x => \delta_x \text{ or } \mu_y => \mu_x + \delta x \tag{21}$$

[9] Note: Type I error exists if H0 is rejected when H0 is true in the state of nature.
Type II error exists if H0 is accepted when H1 is true in the state of nature.

H1: $\mu_y - \mu_x < \delta x$ or $\mu_y < \mu_x + \delta x$ (favorable validation condition) (22)

The critical rejection region for H0 is similarly defined by,

$$T \leq - t(\alpha;\upsilon) \tag{23}$$

except that $\overline{Y}$ and $\overline{X}$ are reversed in the formulation of T such that,

$$\frac{\overline{Y} - \overline{X} - \delta x}{s_{\overline{X}-\overline{Y}}} \leq - t(\alpha;\upsilon) \tag{24}$$

where $\overline{Y}$ - $\overline{X}$ is positive for $\overline{Y} > \overline{X}$.

Note that in this case for $\mu_y > \mu_x$, $\overline{Y}$ - $\overline{X}$ replaces $\overline{X}$ - $\overline{Y}$ in the otherwise identical formulation for $\mu x > \mu y$. Therefore, the test may be written for both conditions of $\overline{Y} > \overline{X}$ and $\overline{X} > \overline{Y}$ using the absolute difference between sample averages, $|\overline{X} - \overline{Y}|$.

$$\frac{|\overline{X} - \overline{Y}| - \delta x}{s_{\overline{X}-\overline{Y}}} \leq - t(\alpha;\upsilon) \qquad \text{for all } \overline{X} \text{ and } \overline{Y} \tag{25}$$

The above formulation with an absolute value on the difference $\overline{X}$ - $\overline{Y}$ for the critical region t-test has the advantage that it reduces the number of t-statistic checks from two to one.

5.1.2 Statistical Formulations of $s_{\overline{X}-\overline{Y}}$ and t($\alpha;\upsilon$)

The degrees of freedom of the Student's t-test statistic, υ, and the form of the equivalent standard deviation are subject to the number of individual measurements, n, used in determining $\overline{X}$ and $\overline{Y}$.

For the assumption that the number of measurements of n_x and n_y tend to infinity, the Student's t-distribution is equivalent to the Gaussian normal distribution. This reaching assumption would allow the use of a simplified normal test statistic, z(α).

Table 5.1 shows the formulation of $s_{\overline{X}-\overline{Y}}$ and t($\alpha;\upsilon$) for three different sets of underlying assumptions, termed Method A, B and C, respectively. For all three methods the form of the equivalent standard deviation is the same:

$$s_{\overline{X}-\overline{Y}} = \sqrt{\frac{s_x^{\,2}}{n_x} + \frac{s_y^{\,2}}{n_y}} \tag{26}$$

The only remaining difference is the choice of the test statistic (z vs. t) and the degrees of freedom in its selection.

The appropriate values for $t(\alpha;\upsilon)$ and $z(\alpha)$ may be found in common reference tables for the Students t-distribution and Gaussian normal distribution in texts on basic statistics [15,16]. Note for the t-statistic that the degrees of freedom may not result in an integer value. Hence, the computed real value should be rounded up to next integer value found in the reference tables [17].

Clearly, Table 5.1 demonstrates that the choice of assumptions reflects onto the calculation complexity of the test statistic of equation (25).

The question toward the end of simplification arises: How much do the test statistic and resulting test outcome vary when each formulation method of Tabele 5.1 is applied to real data? The author addresses these questions in the following Section 5.5.

5.1.3 Alternate Interpretations of the Hypothesis Test

The author presented the hypothesis tests to reflect the forms which are common in basic statistical literature. The hypothesis test may be interpreted in several different fashions:
The difference between the true measurement population averages, μ_x and μ_y, which are estimated by $\overline{X}$ and $\overline{Y}$, is less than the half-tolerance, δx, to a confidence level of (1-α)100 %, if the critical rejection region on H0 is confirmed e.g. if $\dfrac{|\overline{X}-\overline{Y}|-\delta x}{s_{\overline{X}-\overline{Y}}} \leq -t(\alpha;\upsilon)$.

The reader may find that an alternate interpretation better suits the sequence of calibration and validation procedures:
The true measurement average μ_y, estimated by $\overline{Y}$, is within the acceptable tolerance band defined by μ_x +/- dx, where μ_x is estimated by $\overline{X}$, to a confidence level of (1-α)100 % if the critical rejection region on H0 is confirmed e.g. if $\dfrac{|\overline{X}-\overline{Y}|-\delta x}{s_{\overline{X}-\overline{Y}}} \leq -t(\alpha;\upsilon)$.

A third interpretation of the hypothesis tests for validating the true averages as 'equal' is derived as follows:
The true averages, estimated by $\overline{X}$ and $\overline{Y}$, are 'equal' within a maximum difference of δx to a confidence level of (1-α)100 %, if the critical rejection region on H0 is confirmed e.g. if $\dfrac{|\overline{X}-\overline{Y}|-\delta x}{s_{\overline{X}-\overline{Y}}} \leq -t(\alpha;\upsilon)$.

This third interpretation reflects the objective of the validation process for the dependent system validation. In the dependent system Y validation, the average of $\overline{Y}$ is compared to the block average previously defined by tester system X in the favorable expectation of equating calibration conditions.

Table 5.1:
Formulation Methods for the Hypothesis Test Statistic of the Calibration Capability Index[10]

Calibration Capability Formulation Method	METHOD A	METHOD B	METHOD C
Hypothesis Test Name	Smith-Satterthwaite/ Aspin-Welch Test	Small Sample t-Test	Normal Z-test
Null Hypothesis Alternate Hypothesis	H0: $\mu_x - \mu_y \geq \delta_x$ H1: $\mu_x - \mu_y < \delta_x$	H0: $\mu_x - \mu_y \geq \delta_x$ H1: $\mu_x - \mu_y < \delta_x$	H0: $\mu_x - \mu_y \geq \delta_x$ H1: $\mu_x - \mu_y < \delta_x$
Assumptions	• independent X and Y • normal random variables • σ_x and σ_y unknown (estimated by s_x and s_y) • σ_x and σ_y not equal	• independent X and Y • normal random variables • σ_x and σ_y known ($\sigma_x = s_x$ and $\sigma_y = s_y$) • σ_x and σ_y not equal • small n_x and n_y	• independent X and Y • normal random variables • σ_x and σ_y known ($\sigma_x = s_x$ and $\sigma_y = s_y$) • σ_x and σ_y not equal • large n_x and n_y
Test Statistic and Critical Value	Student's t: $t \leq -t(\alpha;\upsilon)$	Student's t: $t \leq -t(\alpha;\upsilon)$	Normal z: $z \leq -z(\alpha)$
Degrees of Freedom v	$v = \dfrac{\left[\dfrac{s_x^2}{n_x} + \dfrac{s_y^2}{n_y}\right]^2}{\dfrac{\left(\dfrac{s_x^2}{n_x}\right)^2}{n_x - 1} + \dfrac{\left(\dfrac{s_y^2}{n_y}\right)^2}{n_y - 1}}$ * rounded up to integer value	$v = n_x + n_y - 2$	$v = \infty$
Equivalent Standard Deviation $s_{\bar{X}-\bar{Y}}$	$\sqrt{\dfrac{s_x^2}{n_x} + \dfrac{s_y^2}{n_y}}$	$\sqrt{\dfrac{s_x^2}{n_x} + \dfrac{s_y^2}{n_y}}$	$\sqrt{\dfrac{s_x^2}{n_x} + \dfrac{s_y^2}{n_y}}$
Ease of Use	Difficult (extra calcs. & tables)	Moderate (tables)	Easy
References	16, 17	18	16

[10] The common method of pooling variances is not an option for this application, since σ_x is not likely to equal σ_y.

5.2 The Development of the Calibration Capability Index, Cc

The Cc index is derived from the formulation of the hypothesis test developed in 5.1:

Eqn. (25): $$\frac{|\bar{X}-\bar{Y}|-\delta x}{s_{\bar{X}-\bar{Y}}} \leq -\,t(\alpha;\upsilon)$$

Multiplying both sides by -1 results in flipping the comparator,

$$\frac{\delta x-|\bar{X}-\bar{Y}|}{s_{\bar{X}-\bar{Y}}} \geq t(\alpha;\upsilon)$$

Dividing both sides by $t(\alpha;\upsilon)$,

$$\frac{\delta x-|\bar{X}-\bar{Y}|}{t(\alpha;\upsilon)\cdot s_{\bar{X}-\bar{Y}}} \geq t(\alpha;\upsilon)/t(\alpha;\upsilon)$$

Therefore,

$$\frac{\delta x-|\bar{X}-\bar{Y}|}{t(\alpha;\upsilon)\cdot s_{\bar{X}-\bar{Y}}} \geq 1 \tag{27}$$

Hence, for the favorable validation condition of $|\mu_x-\mu_y| < \delta x$, the above relationship of equation (27) must hold true.

The author defines the Calibration Capability Index as the left-side of the equation (27):

$$\text{Cc} = \frac{\delta x-|\bar{X}-\bar{Y}|}{t(\alpha;\upsilon)\cdot s_{\bar{X}-\bar{Y}}} \tag{28}$$

Substituting for the common equivalent standard deviation from Table 5.1 yields the expression and critical value for the Calibration Capability Index ,

$$\text{Cc} = \frac{\delta x-|\bar{X}-\bar{Y}|}{t(\alpha;\upsilon)\cdot\left[\frac{s_X^{\,2}}{n_X}+\frac{s_Y^{\,2}}{n_Y}\right]^{1/2}} \geq 1 \tag{29}$$

From the development of the hypothesis test, the limiting condition of calibration capability thus occurs when Cc = 1. The capability of the global system of Fig. 3-1 is therefore increasing as Cc exceeds 1. Increasing capability above 1 improves the confidence for the alternate hypothesis, H1: μ_x-$\mu_y < \delta x$, beyond $(1-\alpha)100\%$.

Note that the equivalent standard deviation in the denominator is equivalent to the component *model for variation of the reference standard* of Equation (15) after applying the Central Limit Theorem.

5.3 Cc measures bias and precision errors

The Cc index includes the difference in $\overline{X}$ and $\overline{Y}$ from measurement data taken in a snapshot in time. The difference between $\overline{X}$ and $\overline{Y}$ may be viewed as a snapshot estimate of the bias error, μ_x-μ_y, between the two tester systems [1]. It is only an estimate of the bias error since the values of the μ_x and μ_y are not actually known.

The denominator in turn is an estimate measure of the potential precision error between the two systems. Again, the sample standard deviations, s_x and s_y are only estimates of the true standard deviations, σ_x and σ_y. Therefore, the Cc index is a cumulative, but local measure of both bias and precision errors in the nominal average hardness between the two reference systems.

Estimates of μ_x and μ_y, as well as σ_x and σ_y, improve with increasing number of measurements. Their final best-estimate values are determined to the highest degree of confidence when the tester system has placed its maximum allowable number of measurements on the reference block with roughly 120 to 150 measurements.

5.4 Development of the Zeroed Calibration Capability, Cc0

Because the measurement averages $\overline{X}$ and $\overline{Y}$ are estimates of the μ_x and μ_y, it is possible that the true block means, μ_x and μ_y, are not actually within the band of the half-tolerance *even though* $\overline{X}$ and $\overline{Y}$ are momentarily measured to be identical e.g. the absolute difference $\overline{X}-\overline{Y}$ goes to zero. Although the reader may intuitively perceive such a condition to be odd, the probability that the two means are not within δx at $(1-\alpha)100\%$ of the time may result due to the large variation on both $\overline{X}$ and $\overline{Y}$ (e.g. s_x and s_y large). That is, it was just momentary chance that $\overline{X}$ and $\overline{Y}$ were identical in the snapshot when the measurements were taken.

When μ_x and μ_y are not identical, such a bias error may be attributed to human error in calibrating the dependent tester system to the nominal average hardness of the block. Large variation in the distributions of X and Y will cause the operators to make a faulty initial calibrations.

It might also be the case that μ_x and μ_y, the centers of their respective block measurement distributions, are indeed equal. However, the large variation spread in their distributions may cause future snapshots in time to yield subgroup averages $\overline{Xi}$ and $\overline{Yi}$'s that differ by more than the half-tolerance band.

It may therefore be of interest to remove the contribution of bias error from the Calibration Capability index to attain only relative measure of the potential precision error.

The author terms this treatment for $\bar{X} - \bar{Y} = 0$ the *Zeroed Calibration Capability, Cc0.* It is defined as,

$$\text{Cc0} = \frac{\delta x}{t(\alpha;\upsilon)\cdot\left[\frac{s_x^{\,2}}{n_x}+\frac{s_y^{\,2}}{n_y}\right]^{1/2}} \geq 1 \qquad (30)$$

The term 'zeroed' reflects the condition of calibration verification where the measurement average of the dependent system Y is adjusted to be identical to the measurement average of the parent system X, such that their nominal calibration points are 'zeroed'.

A Cc0 of 1 reflects the likelihood of approximately α % that the average $\bar{Y}'$ of the subsequent sample group will be outside of the allowable half-tolerance band, even though the testers were calibrated to be identical with equal true averages $\mu_x = \mu_y$. Note that $\bar{Y}'$ is from the same population as $\bar{Y}$ defined by the block and the dependent system.

The Cc0 is an improved metric for measuring only the precision error resulting from the combination of two local tester systems. It is also a measure of the 'best possible' Calibration Capability achievable assuming no mean offset, e.g. $\bar{X} - \bar{Y} = 0$.

5.5 Referencing Measurement Quantity

The significant influence of measurement quantity on Calibration Capability suggests that for practical application or comparison between tester systems the Calibration Capability index should be stated on the basis of the number of measurements. The index includes two subgroups of n_x and n_y measurement quantities. The author suggests that these be included in the index terminology to allow for equivalent comparisons between systems.

The index terminology must account for varying quantities of measurements performed in calibration of the standardized blocks, as well as to the discretion of the user in validation.

The author proposes the following terminology for the Calibration Capability index to make specific reference to the measurement quantities:

$$\text{Cc}^{\,nx}_{\quad ny}$$

For example, C^6_5 represents $n_x = 6$ measurements performed in block calibration for the parent system and $n_y = 5$ by user tester validation of the dependent system.

5.6 Referencing Hardness Level

It was demonstrated in Chapter 2.0 that the measured hardness variance tends to decrease for increasing hardness using commercial tester technologies over the spectrum of hardnesses levels from HRC 25 (soft) to HRC 63 (hard). Refer to Figure 2.6. Hence, to draw meaningful comparisons between global measurement systems only Cc indexes computed at equivalent nominal hardness levels should be compared. In order to ensure that global systems of different nominal hardness levels are not falsely or inadvertently compared, the author recommends the following reference to the nominal hardness level in the index terminology:

$$Cc(\text{scale/nominal H-value})^{nx}{}_{ny}$$

For example, $Cc(C45)^{6}{}_{5}$ for the Calibration Capability at HRC 45 nominal hardness. Note that nominal hardness levels are specified in ASTM E-18 [4].

5.7 Interpretation of Calibration Capability

The Calibration Capability allows for the determination of the following undesirable conditions. In these conditions the total calibration system consisting of two tester systems and a common reference block is not 'capable' of satisfying the hardness half-tolerance specification by the manufacturer. These conditions describe the relationship of the averages of two measurement subgroups taken on the same test block:

Conditions of Non-Calibration Capability

Condition 1: The two measurement systems are off-center from nominal at a difference greater than the block half-tolerance e.g. $|\,\bar{X} - \bar{Y}\,| > \delta x$

Condition 2: The local system variation for the measurement averages of system X and system Y is too large relative to the specified half-tolerance. e.g. s_x and/or s_y are large or n_x and/or n_y are small.

Condition 3: The two local measurement systems X and Y are off-center and have large variation.

The reader may better understand the respective conditions by studying graphical representations of each. Note that the bell-shaped curve for the normal distribution of *subgroup averages* was previously validated in Chapter 3. By inspection of Figures 5-1, 5-2 and 5-3 the reader will attain an improved understanding of how the Calibration Capability gages the probability that subgroup averages $\bar{X}$ and $\bar{Y}$ will be equal within the half-tolerance.

The limiting condition 2 of Figure 5-2 shows the condition of Zeroed Calibration Capability. Although the two test systems are centered at the identical true average, they may demonstrate a difference in measurement averages that falls outside of the half-tolerance. This condition stems from excessively large variation in one or both of the test systems (relative to the tolerance). As a result of Condition 2, a user may be falsely inclined to adjust the mean setting of their tester since $\bar{X}$ and $\bar{Y}$.

Figure 5-1:
Condition 1: The two measurement systems are off-center at a difference greater than δx.

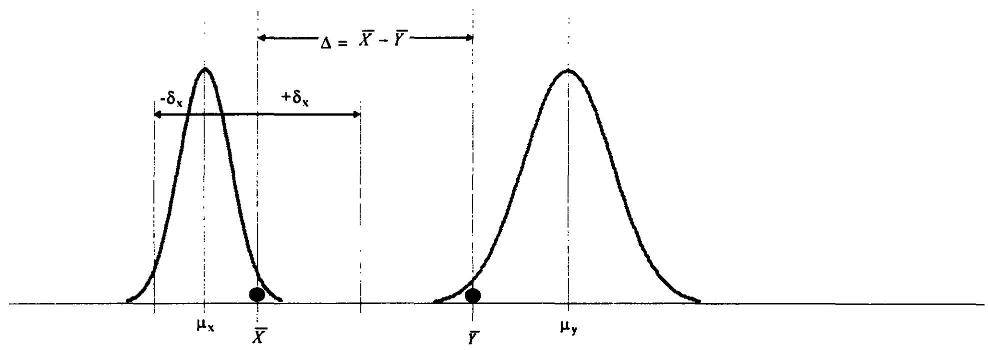


Figure 5-2:
Condition 2: The local system variation is too large relative to the specified half-tolerance, δx.

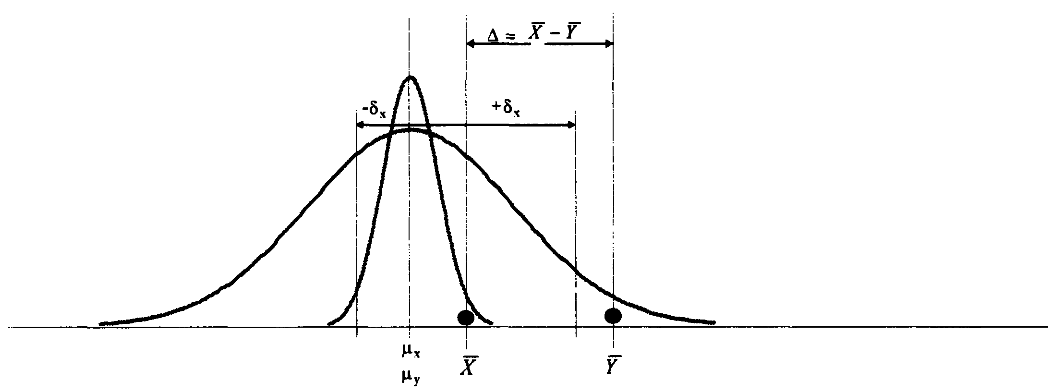
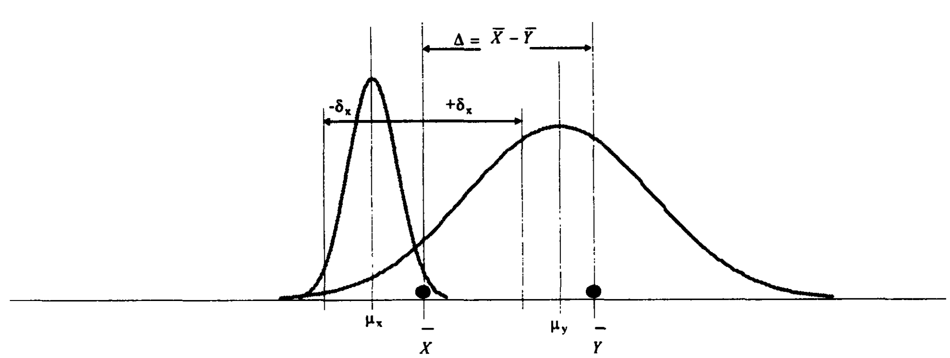
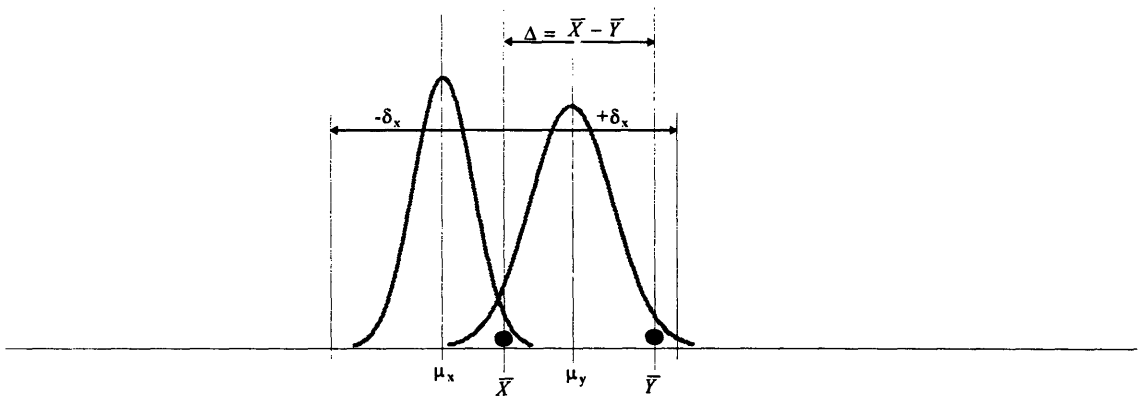


Figure 5-3:
Condition 3: The two local measurement systems X and Y are off-center *and* have large variation.



The favorable condition of calibration capability for the global system is depicted in Figure 5-4 below. Note that the capability is shown to be marginal as the distribution of $\bar{Y}$ has a small tail that extends beyond the half-tolerance, δx.

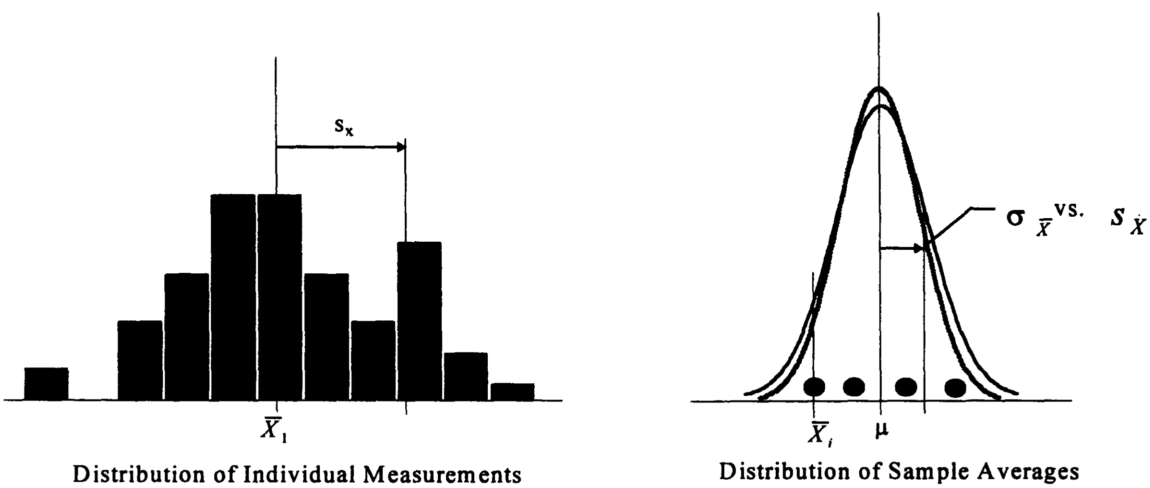
Figure 5-4: Condition of (marginal) calibration capability (Cc > 1).



5.8 Illustration of the Effect of Sampling Error in Quantifying Variation

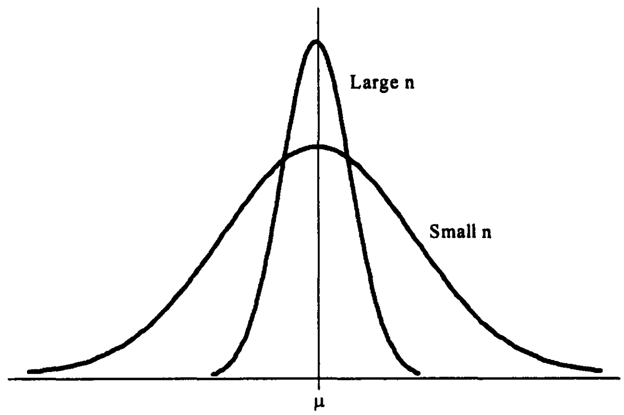
The reader should note a key assumption underlying the simplified graphical: The inherent standard deviation of the block, σ_x, equals the sample standard deviation, s_x. Due to the small number of measurement data points, the sample standard deviation represents an *estimate* of the underlying true standard deviation for large n. Therefore, the standard deviation of the subgroup averages $s_{\bar{X}}$ through conversion by the Central Limit Theorem of equation (14) also includes this estimation error. As a result, the 'true' shape and width of the distribution for subgroup averages may deviate slightly, as depicted in Figure 5-5 below.

Figure 5-5: Estimation of $\sigma_{\bar{X}}$ by $s_{\bar{X}}$



Not only does the estimate of $\sigma_{\bar{X}}$ improve with increasing number of subgroup measurements, but $\sigma_{\bar{X}}$ also decreases with increasing n by the C.L.T. equation of equation (14). The bell curve of the subgroup averages therefore becomes narrower with increasing n. Refer to Figure 5-6 below.

Figure 5-6: The reduction of standard deviation for subgroup averages, $s_{\bar{X}}$, with increasing subgroup size.



In more practical terms, the subgroup average is more likely to be closer to the grand average of the block with increasing number of measurements. In the limit as n goes to its maximum, the subgroup average equals the grand average. The Calibration Capability reflects this property of the Central Limit Theorem as increasing number of subgroup measurements improves the likelihood that the user can validate their tester system within the prescribed tolerance engraved on the block without (unnecessary) adjustment to their system.

5.9 Interpretation of the Calibration Capability using the Distribution of $\bar{X}$ - $\bar{Y}$

The previous graphical representations are intended to assist the reader who is new to statistical thinking. A description that better reflects the development of the hypothesis test on μ_x-μ_y (on which the Cc index is based) is one that characterizes the Conditions of Non-Capability in terms of the distributions of $\bar{X}-\bar{Y}$. It should be noted that on the assumption that $\bar{X}$ and $\bar{Y}$ are normal, then so will their difference (or sum) be normal.

Figure 5-7 depicts *Condition 1* in which the distribution of the true mean difference, μ_x-μ_y, is centered far beyond the half-tolerance band $\pm$ δx about 0. As a result, any sample difference of $\bar{X}$ - $\bar{Y}$ is likely to be out of the tolerance specification.

Figure 5-8 represents *Condition 2* in which the true mean difference μ_x-μ_y is centered at 0. However, because of the large variation in $\bar{X}-\bar{Y}$ the sample difference $\bar{X}-\bar{Y}$ is probable of falling outside the half-tolerance band. The large variation may be due to a large standard deviation $s_{\bar{X}}$ or $s_{\bar{Y}}$ in either the X or Y sample or due to a small number of measurements.

Figure 5-9 shows the combined *Condition 3* in which the true mean difference μ_x-μ_y is offset and the variation is large. The large area under the distribution curve that is outside the half-tolerance band represents a high probability that the subsequent sample differences $\overline{X}-\overline{Y}$ will fall outside of the half-tolerance band about 0.

Figure 5-7: *Condition 1* where $\mu_x >> \mu_y$

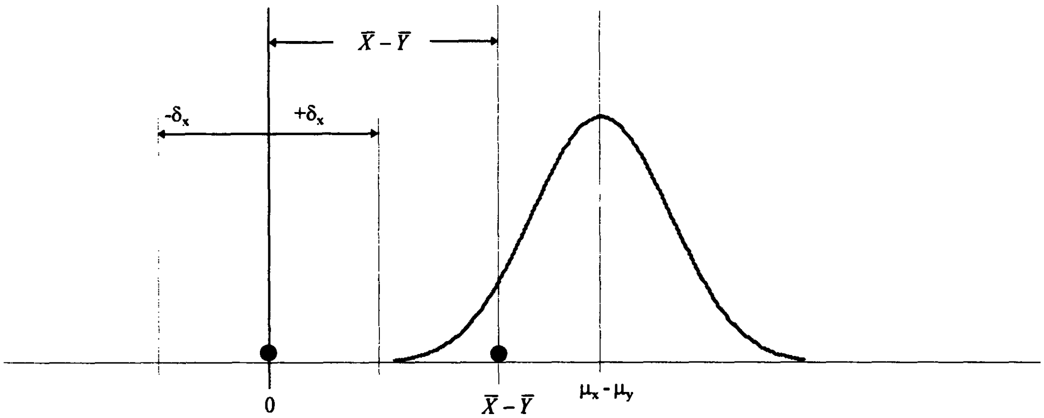


Figure 5-8: *Condition 2* where $\mu_x = \mu_y$, but the variation $s_{\overline{X}-\overline{Y}}$ is large.

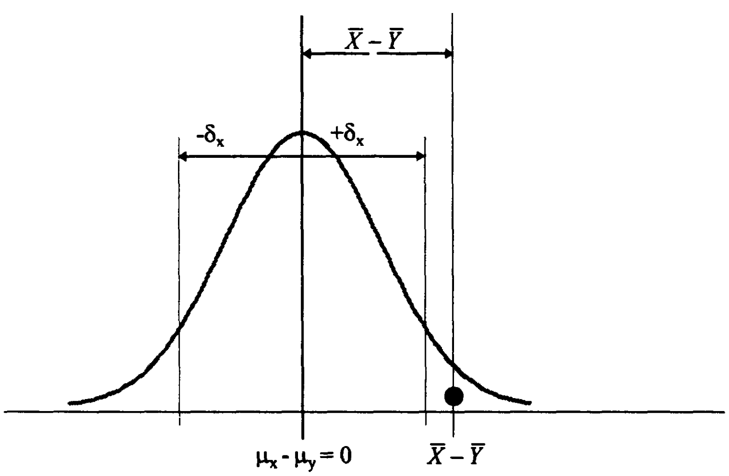
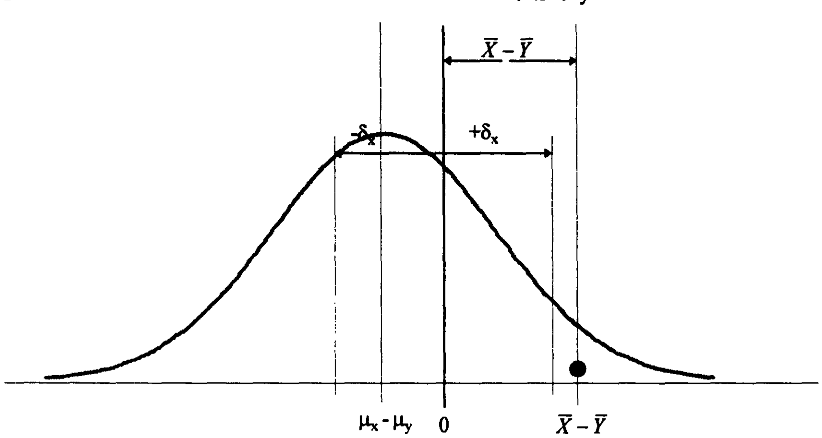


Figure 5-9: Combined *Condition 3* where μ_x -μ_y is offset and the variation $s_{\overline{X}-\overline{Y}}$ is large



5.10 Practical Definitions of Calibration Capability

What does it mean when the global calibration system is capable ? The author presents the following three viewpoints that support comprehension and applicability of the Calibration Capability index:

One interpretation is that the half-tolerance is sized large enough in relation to the variation of the manufacturer's and user's tester systems (including that variation attributed to the block) that it can be reliably expected that a small difference in measured averages will not exceed the half-tolerance.

Conversely, the combined variation of the manufacturer and user test systems (including the block variation) is small enough so that the user can reliably calibrate within the hardness range specified by the manufacturer (engraved on the block) even though there exists some offset in true means between the test systems. After all, calibration by sampling a block of variable hardness in tester production precludes that there will be some offset even without tester usage drift (wear).

Finally, the author applies the perspective of the Central Limit Theorem: The quantity of measurements taken on the same block in both the parent and dependent system are large enough so that the effects of tester variation are reduced. In addition, the resulting improved estimates of their local grand averages result in a smaller possible difference in subgroup averages. The differences are thus more likely to lie within the specified half-tolerance.

5.11 A Simulation Study of the Calibration Capability

A simulation was conducted using real measurement data from two tester systems on seven common blocks of two types, Regular and Large, over a variety of HRC hardness levels and measurement quantities. The Instron 600S was used to represent the parent system X calibrating the block; the Instron 500S represented the dependent system Y of the customer environment.

The objectives for the simulation are to:

- evaluate the Calibration Capability indices of the 3 candidate methods (A, B and C) of Table 5.1 for the test statistics $t(\alpha;\upsilon)$. This comparison demonstrates the effect of loosening the underlying assumptions, while assessing the penalties for simplification to the normal Z-test of Method C.
- study the effects of varying sample sizes/measurement quantities on calibration capability.
- investigate the system effects of combining measurement subgroups of high variation with low variation, as well as joining high sample sizes with low sample sizes.

The original measurement data sets consisting of 30 measurement per block by each tester system may be found in Appendix C. The half-tolerances applied in the study are those specified by ASTM E-18 as the +/- tolerance for test blocks for the respective hardness level (+/- 0.5 for HRC 60 and above, +/- 1.0 for below HRC 60). Refer to Table 1.1 of the Introduction in Chapter 1.

The significance level applied for the test statistics was chosen as $\alpha = 0.005$, corresponding to an equivalent confidence of $(1-\alpha)100\% = 99.5\%$. Note that the z-statistic for the normal test of Method C uses a z-value of 3, representing the common 3-sigma level. True conversion of $z=3$ yields a significance of $\alpha=0.0025$ or 99.75% confidence level for the one-tail test.

Note that α is used here as a measure of relative risk in having the two tester calibration points (averages) differ. The effect of relaxation of the underlying assumptions is deemed to outweigh this nuance in a 0.25 % difference in significance level.

The simulation results are tabulated in Table 5.2.

5.11.1 Results from the Cc Simulation Study

It is apparent that the sample standard deviations from the same local tester system may change dramatically over the varying measurement quantities. For instance, at $n = 6$, $s_x = 0.343$ vs. 0.282 at $n = 30$. The same HRC25 block with a subsequent subgroup of $n = 6$ demonstrates a dramatically lower $s_x = 0.148$. Thus the fact that the sample standard deviations are only estimates of the true standard deviation (of maximized n) is reflected in the inconsistent nature of the simulation data. This is also clearly a violation of the assumption of Method C that assumes that the true population standard deviations are known.

In turn, the differences in measurement averages from a common block are also not consistent. The notion that both the averages and standard deviations are probabilisticaly distributed and that their sample values reflect estimates of the distribution grand average at a snapshot in time is thus supported by the simulation data.

As expected the system sets with n_x and n_y of 30 data points have the lowest equivalent standard deviation and hence the highest Cc values..

In comparing the t-test statistics between method A and method B, it is apparent that the degrees of freedom for the Aspin-Welch test (Method A) are never the highest and are generally lowest when compared to the other two methods. Method B thus has higher t-values than Method A.

Note that the applied values for the t-statistics resulted as follows for the three methods over all hardnesses:

Method A:	Average $t = 3.25$	6-5 n-combination $t = 3.25$ to 4.03
Method B:	Average $t = 3.01$	6-5 n-combination $t = 3.25$

In general, Method C using the test statistic $z=3$ yields calibration capabilities that are higher. The average and maximum differences in Cc between Methods A and C are 11.5% and 25.6%, respectively.

One may therefore conclude that the use of z-statistic, $z=3$ in relaxing the underlying assumptions results in slightly inflated calibration capabilities.

Table 5.2: Simulation of Calibration Capability using 600/500 Global System

		Parent 600 System					*Dependent 500 System*								*Method A*					*Method B*				*Method C*: z = 3	
HRC Nom.	Block Serial No.	nx	sx	Xbar	Rx	95% C.I.	ny	sy	Ybar	Ry	sx/sy	Equiv. s	Xbar-Ybar	half-tolerance dx	V	Vint.	t(.005,V)	*CcA*	*Cc0A* for Xbar-Ybar = 0	V	t(.005,V)	*CcB*	*Cc0B* for Xbar-Ybar = 0	*Ccc*	*Cc0c* for Xbar-Ybar = 0
Regular Grade																									
25	H00128	6	0.343	26.27	1.05	0.46	5	0.164	26.12	0.40	2.09	0.158	0.150	1.0	7.4	8	3.355	1.60	1.89	9	3.250	1.65	1.95	1.79	2.11
		10	0.282	26.26	1.05	0.27	5	0.164	26.12	0.40	1.72	0.115	0.140	1.0	12.5	13	3.012	2.47	2.88	13	3.012	2.47	2.88	2.48	2.89
		10	0.282	26.26	1.05	0.26	10	0.313	26.34	1.00	0.90	0.133	-0.080	1.0	17.8	18	2.878	2.40	2.61	18	2.878	2.40	2.61	2.30	2.50
		30	0.224	26.24	1.12	0.11	30	0.255	26.38	1.00	0.88	0.062	-0.140	1.0	57.1	58	2.663	5.21	6.06	58	2.663	5.21	6.06	4.63	5.38
		6	0.148	26.19	0.38	0.20	5	0.313	26.56	0.80	0.47	0.152	-0.370	1.0	5.5	6	3.707	1.11	1.77	9	3.250	1.27	2.02	1.38	2.19
45	G00390	6	0.145	46.04	0.40	0.19	5	0.152	45.66	0.40	0.95	0.090	0.380	1.0	8.5	9	3.250	2.12	3.41	9	3.250	2.12	3.41	2.29	3.70
		10	0.139	46.04	0.40	0.13	5	0.152	45.66	0.40	0.91	0.081	0.380	1.0	7.5	8	3.355	2.28	3.68	13	3.012	2.54	4.10	2.55	4.12
		10	0.139	46.04	0.40	0.13	10	0.197	45.81	0.60	0.71	0.076	0.230	1.0	16.2	17	2.898	3.48	4.53	18	2.878	3.51	4.56	3.37	4.37
		30	0.135	46.03	0.47	0.07	30	0.200	45.93	0.70	0.68	0.044	0.100	1.0	50.9	51	2.676	7.63	8.48	58	2.663	7.67	8.52	6.81	7.57
		6	0.139	46.02	0.38	0.18	5	0.192	45.92	0.50	0.72	0.103	0.100	1.0	7.2	8	3.355	2.61	2.90	9	3.250	2.69	2.99	2.91	3.24
63	R02539	6	0.063	63.72	0.16	0.08	5	0.192	63.48	0.50	0.33	0.090	0.240	0.5	4.7	5	4.032	0.72	1.38	9	3.250	0.89	1.72	0.97	1.86
		10	0.055	63.70	0.16	0.05	5	0.192	63.48	0.50	0.29	0.088	0.220	0.5	4.3	5	4.032	0.79	1.42	13	3.012	1.06	1.89	1.07	1.90
		10	0.055	63.70	0.16	0.05	10	0.133	63.50	0.50	0.41	0.046	0.200	0.5	12.0	12	3.055	2.16	3.60	18	2.878	2.29	3.82	2.20	3.66
		30	0.078	63.70	0.40	0.04	30	0.123	63.51	0.50	0.63	0.027	0.190	0.5	49.1	50	2.678	4.35	7.02	58	2.663	4.38	7.06	3.89	6.27
		6	0.119	63.74	0.36	0.16	5	0.122	63.50	0.30	0.98	0.073	0.240	0.5	8.6	9	3.250	1.10	2.11	9	3.250	1.10	2.11	1.19	2.28
Large Grade																									
30	95I30005	6	0.211	30.23	0.55	0.28	5	0.152	30.26	0.30	1.39	0.110	-0.030	1.0	8.9	9	3.250	2.72	2.80	9	3.250	2.72	2.80	2.95	3.04
		10	0.212	30.29	0.59	0.20	5	0.152	30.26	0.30	1.39	0.095	0.030	1.0	11.0	11	3.106	3.27	3.37	13	3.012	3.37	3.48	3.39	3.49
		10	0.212	30.29	0.59	0.19	10	0.145	30.21	0.40	1.46	0.081	0.080	1.0	15.9	16	2.921	3.88	4.22	18	2.878	3.94	4.28	3.78	4.10
		30	0.161	30.30	0.59	0.08	30	0.170	30.29	0.70	0.95	0.043	0.010	1.0	57.8	58	2.663	8.70	8.78	58	2.663	8.70	8.78	7.72	7.80
		6	0.123	30.24	0.30	0.16	5	0.239	30.32	0.60	0.51	0.118	-0.080	1.0	5.7	6	3.707	2.10	2.28	9	3.250	2.40	2.61	2.60	2.82
40	95I40004	6	0.092	40.50	0.21	0.12	5	0.114	40.34	0.30	0.81	0.063	0.160	1.0	7.7	8	3.355	3.95	4.71	9	3.250	4.08	4.86	4.42	5.26
		10	0.081	40.51	0.21	0.08	5	0.114	40.34	0.30	0.71	0.057	0.170	1.0	6.1	7	3.499	4.16	5.01	13	3.012	4.83	5.82	4.85	5.84
		10	0.081	40.51	0.21	0.07	10	0.116	40.33	0.30	0.70	0.045	0.180	1.0	16.1	17	2.898	6.32	7.71	18	2.878	6.37	7.77	6.11	7.45
		30	0.091	40.52	0.34	0.04	30	0.141	40.35	0.60	0.65	0.031	0.170	1.0	49.6	50	2.678	10.12	12.19	58	2.663	10.17	12.25	9.03	10.88
		6	0.125	40.51	0.33	0.17	5	0.179	40.32	0.40	0.70	0.095	0.190	1.0	7.0	7	3.499	2.44	3.01	9	3.250	2.63	3.24	2.84	3.51
50	95I50005	6	0.068	49.90	0.20	0.09	5	0.179	49.52	0.40	0.38	0.085	0.380	1.0	5.0	5	4.032	1.81	2.93	9	3.250	2.25	3.63	2.44	3.93
		10	0.070	49.92	0.25	0.07	5	0.179	49.52	0.40	0.39	0.083	0.400	1.0	4.6	5	4.032	1.79	2.99	13	3.012	2.40	4.00	2.41	4.01
		10	0.070	49.92	0.25	0.06	10	0.149	49.57	0.50	0.47	0.052	0.350	1.0	12.8	13	3.012	4.15	6.38	18	2.878	4.34	6.67	4.16	6.40
		30	0.060	49.94	0.25	0.03	30	0.120	49.59	0.50	0.50	0.024	0.350	1.0	42.6	43	2.695	9.85	15.15	58	2.663	9.96	15.33	8.85	13.61
		6	0.050	49.96	0.13	0.07	5	0.055	49.56	0.10	0.91	0.032	0.400	1.0	8.3	9	3.250	5.78	9.63	9	3.250	5.78	9.63	6.26	10.43
60	95I60001	6	0.089	60.44	0.26	0.12	5	0.167	60.06	0.40	0.53	0.083	0.380	0.5	5.9	6	3.707	0.39	1.62	9	3.250	0.44	1.85	0.48	2.01
		10	0.080	60.43	0.26	0.08	5	0.167	60.06	0.40	0.48	0.079	0.370	0.5	4.9	5	4.032	0.41	1.57	13	3.012	0.55	2.11	0.55	2.11
		10	0.080	60.43	0.26	0.07	10	0.185	60.09	0.70	0.43	0.064	0.340	0.5	12.3	13	3.012	0.83	2.60	18	2.878	0.87	2.73	0.84	2.61
		30	0.076	60.40	0.35	0.04	30	0.127	60.09	0.70	0.60	0.027	0.310	0.5	47.4	48	2.682	2.62	6.90	58	2.663	2.64	6.95	2.34	6.17
		6	0.101	60.39	0.29	0.13	5	0.045	60.12	0.10	2.24	0.046	0.270	0.5	7.2	8	3.355	1.49	3.25	9	3.250	1.54	3.35	1.67	3.63

By inspecting the standard t-values on a table for the Student's t-distribution for the global system of $n_x = 6$ and $n_y = 5$ ($\upsilon = 5+6-2 = 9$), an $\alpha = 0.01$ still yields a t(0.01;9) of 2.821.

Therefore, we may conclude that the Method C using a normal test statistic of $z = 3$ is more than adequate for use in a relative capability measure for varying sample sizes with converted confidence levels greater than 98% instead of 99.5%+.

The effects of increasing n_x

The simulation of Table 5.2 demonstrates that increasing the parent system's measurement quantity n_x from 6 to 10 does result in the expected increase in Cc. However, the increase in measurement quantity alone is not a very high leverage for Cc improvement. This is particularly the case at the higher hardness levels of HRC 50 and above. The main reason for this is that at the higher hardnesses the dependent system's variance (s_y^2) is large relative to that of the parent system, as shown in the table (refer to sources discussion). In general, the parent system is often at a state of higher technology and is operated and maintained in a more controlled fashion. As a result, the variance of the dependent system (s_y^2) dominates the equivalent variance. Therefore, increasing n_x yields improvements with diminishing returns.

The effects of increasing n_y

It is also shown by the simulation that increasing the dependent system's measurement quantity n_y to 10 does yield significant improvements in Cc. The Cc index can be used by both users (dependent systems) and manufacturers (parent systems) to evaluate the benefits against the costs of increasing the users measurement quantity from 5 to 10.

Examples of Cc less than 1 due to mean offsets

The Cc results for the HRC 60 block simulations demonstrate conditions of non-capability, as Cc is less than 1. This is attributed to the large average difference ($\bar{X} - \bar{Y}$) and the large standard deviation of the dependent system (s_y), both working against the tighter half-tolerance ($\delta x = 0.5$). The Cc0 values greater than 2 allow us to conclude that a reduction in the average difference $\bar{X}$ - $\bar{Y}$ would make the system capable at all measurement quantities.

In such an instance, it may be tempting to achieve calibration capability by a machine offset adjustment. The other hardness levels show us that a reduced difference $\bar{X} - \bar{Y}$ is feasible. The author cautions about haphazard use of this method. Frequent and unwarranted tester adjustments in reaction to the natural variation (defined by s_y) may result, a behavior Devor et al term 'over-control' [14].

5.11.2 General Conclusions from the Cc Simulation Study

The simulation study yields the following conclusions:

1. *The normal test statistic, $z = 3$, works for small sample sizes.* The Calibration Capability Method C using the Z-test statistic of $z = 3$, assuming a Gaussian normal distribution of hardness measurements is an adequate approximation with respect to the other Methods A and B. Method C still provides effective confidence levels greater than 95% for Cc = 1. The benefits of

simplification, e.g. ease of calculation without the use of statistical tables, outweigh the cost in equivalent significance level for use as a relative metric..

The simplified and final form of the Calibration Capability, Cc is therefore:

$$\mathrm{Cc} = \frac{\delta x - \left|\bar{X} - \bar{Y}\right|}{3 \cdot \left[\frac{s_x^{\,2}}{n_x} + \frac{s_y^{\,2}}{n_y}\right]^{1/2}} \geq 1 \tag{31}$$

and the Zeroed Calibration Capability, Cc0 is thus:

$$\mathrm{Cc0} = \frac{\delta x}{3 \cdot \left[\frac{s_x^{\,2}}{n_x} + \frac{s_y^{\,2}}{n_y}\right]^{1/2}} \geq 1 \tag{32}$$

2. *Perform several permutation trials to calculate Cc.* The sample averages and standard deviations for small subgroups of n less than 10 behave as random estimates of the true block averages and average standard deviations. The measurement subgroups represent snapshots in time that were shown to vary significantly. As a result, the estimates of s_x and s_y will not remain constant over a series of trials or applications.

For the application of the Cc index it is therefore recommended to calculate an *average or minimum* Cc (and Cc0's) index over several trials with measurement data of identical sample sizes. The goal is to achieve an index that is representative of the total variation present. A practical and effective number of trials is proposed as 3 for a Cc^6_5.

3. *Increase the measurement quantity with the system of highest variance.* If the parent's system has a significantly lower variance s_x^2 than the dependent systems s_y^2, increasing the parent measurement quantity n_x alone is a low leverage policy for achieving improvements in calibration capability.

A more effective policy is to increase the measurement quantity on the local system with the highest variance, which is expected to be the dependent system of the users/customers. The benefits and costs of this additional burden on the users may be evaluated using the Calibration Capability with quantifiable trade-offs in confidence by varying the controllable parameters of n_x, n_y, δ_x.

4. *Calibration Capability is feasible with the current state of technology.* The Zeroed Calibration Capabilities (Cc0) of the simulation demonstrate that for a representative 600 and 500 global system, system calibration capability can be achieved with respect to standard half-tolerances in

all cases if the initial mean offset $\bar{X} - \bar{Y}$ can be minimized. Refer to Table 5.2 comparing Cc with the corresponding Cc0 value. The offset may be controlled by Instron prior to shipment or may require more frequent monitoring of the customer tester by the customer. However, any change to either the parent or dependent system, for example the change-out of the indenter, can cause a mean offset with an adverse effect on marginal capabilities.

5. *Use caution in adjusting for mean offsets.* The author also cautions about 'over-control' of the tester systems that may result from relying on machine offsets to improve Cc by reducing the average difference $\bar{X} - \bar{Y}$. Reaction to the natural variation (defined by sy) of the total system by frequent and unwarranted tester adjustments makes the it more difficult to maintain control of the tester's calibration settings. General characterization of the global system using the Calibration Capability requires the tester systems to be in statistical control.

5.12 Comparison to Another Statistical Methodology

NAMAS, the European standardizing body, employs similar statistical techniques in their expression for uncertainty in hardness measurements in specification NIS 0406 [9].
For example, the random uncertainty on the measurement average is determined by multiplying a mean deviation, $s/n^{1/2}$, by the tabulated t-statistic for small measurement subgroups (less than 10).

NAMAS uses a uncertainty (U) value for the inter-comparison of the measurement average to the calibration certificate value [9]. The uncertainty value is a calculation of the mean difference plus/minus an equivalent 95% confidence interval.

This method differs from this author's approach of the Calibration Capability in that:
- it treats the average uncertainty, stated as the 95% C.I. tolerance on the certificate, as known and constant based on a single measurement trial
- it does not provide an acceptance criteria e.g. is a uncertainty value for $\bar{X} - \bar{Y}$ of 0.83 to -0.37 good or bad ?

5.13 Assessing the Returns from Variation Reduction using the Cc0 Index

The (zeroed) calibration capability index is a useful tool to answer to the question: How much of an improvement in total measurement variation is detectable by the user ? What is important to the user is not the amount of variation per se, but rather the ability of his tester to precisely measure the average hardness recorded on the calibration certificate within the allowable tolerance.

The previous 600/500 system simulation of Table 5.2 demonstrated that the use of Large grade blocks results in $Cc0^6{}_5$ values greater than 2 at the C60 level and greater than 3 at the lower hardness levels. The components of variance framework of Chapter 2 showed that a reduction in block variance affects both the variation seen in the parent and dependent systems, s_x^2 and s_y^2. Hence, the denominator of the Cc index is reduced.

$$Cc = \frac{\delta x - |\bar{X} - \bar{Y}|}{3 \cdot \left[\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y} \right]^{1/2}} \geq 1$$

where $s_x^2 = s_{X\ tester}^2 + s_{block}^2$ and $s_y^2 = s_{Y\ tester}^2 + s_{block}^2$

For the example, for the HRC 30 block, 95I30005 (Refer to Table 5.2):

$$(s^2_{block})_{max} = s^2_{DWmeasure} = .067^2 = 0.00449$$

$$s_x^2 = s^2_{block} + s^2_{tester}$$
$$.161^2 = .067^2 + .146^2$$

Similarly,

$$s_y^2 = s^2_{block} + s^2_{tester}$$
$$.170^2 = .067^2 + .156^2$$

Note that the measurement variance terms are dominated by the tester variation.

For $n_x = 6$ and $n_y = 5$ and δx=1.0, the Cc0 index is,

$$Cc0 = \frac{1.0}{3 \cdot \left[\frac{.161^2}{6} + \frac{.170^2}{5} \right]^{1/2}} = \quad \mathbf{3.3}$$

Now let's presume that another 30% improvement is achieved in block variation,

In this case, $s^2_{block} = 0.7 * .067^2 = .056^2$

The measurement variation components are thus computed:

$$s_x^2 = s^2_{block} + s^2_{tester}$$
$$.156^2 = .056^2 + .146^2$$

and

$$s_y^2 = s^2_{block} + s^2_{tester}$$
$$.166^2 = .056^2 + .156^2$$

The resulting Cc0 index is therefore,

$$Cc0 = \frac{1.0}{3 \cdot \left[\frac{.156^2}{6} + \frac{.166^2}{5} \right]^{1/2}} = \quad \mathbf{3.4}$$

For the 30% improvement in block variation, only a 3% improvement in Cc0 was achieved. Thus it is shown that due to the larger contribution of tester variation over block variation, further focus on improving the block (beyond the current Large grade) will yield only marginal returns in calibration capability.

5.14 Simulation of improvements in tester technology

In the previous section it was demonstrated that future improvements in calibration capability are to be attained by improving the variation attributed to the tester. As a reminder, 'tester' represents the combined sources of the machine, the indenter and the operator influence.

A good model of technology enhancement for reduction in machine variation is the deadweight tester operated by NIST. The measurement data for the Large blocks at HRC 30, 40, 50 and 60 can linked to the dependent systems of the 500 and 600 tester for measurements taken on the same blocks.

The simulation using the NIST deadweight tester data to represent the parent system is presented in Table 5.3 in combination with the 600 and 500 dependent systems. The target Cc is established for this simulation as 1.5 in order to compensate for the outliers in hardness measurement and the inferior estimates of the standard deviation at low sample sizes.

Note that the large improvement in tester variation of the parent system for the deadweight tester compared to Table 5.2. The DW/600 and the DW/500 global systems also show a dramatic improvement in Cc0. Note that at the HRC 60 level the Cc index is less than zero because the mean offset, $\bar{X} - \bar{Y}$, is larger than the allowable tolerance.

Both the NIST Deadweight and 600 systems display such low total measurement variances at HRC 60 that they have a significantly higher Cc0 than the DW/500 system at the same hardness level. The Cc0 is almost sufficient enough to compensate for the large mean offsets, 0.325 to 0.377 relative to a 0.5 allowable half-tolerance.

The Cc0 improvement for the DW/500 system at the highest HRC 60 hardness is negligible in comparison to Table 5.2. This is because the 600 tester previously demonstrated very low measurement variation relative the 500 tester system.

5.15 Improving the mean offset vs. reducing variation

The previous simulation demonstrated the trade-off embodied by the calibration capability between reduction in measurement variation and reduction in the mean offset $\bar{X} - \bar{Y}$. The calibration capability allows for the assessment in performance improvement relative to the costs associated with either reduction in mean offset or reduction in global variation. Note that the Calibration Capability is measured between the parent system of the Standards Testers and dependent system of testers supplied to customers.

Table 5.3: Calibration Capability Simulation for Reduction of Parent System Tester Variation

													Method A					Method B				Method C: z = 3	
HRC Nom.	Block Serial No.	nx	sx	Xbar	95% C.I.	ny	sy	Ybar	sx/sy	Equiv. s	Xbar-Ybar	half-tolerance dx	V	Vint.	t(.005,V)	Cc_A	$Cc0_A$ for Xbar-Ybar = 0	V	t(.005,V)	Cc_B	$Cc0_B$ for Xbar-Ybar = 0	Cc_C	$Cc0_C$ for Xbar-Ybar = 0
		Parent Deadweight System				*Dependent 600 System*																	
30	95I30005	6	0.211	30.23	0.27	6	0.052	30.00	4.03	0.089	0.234	1.0	5.6	6	3.707	2.33	3.04	10	3.169	2.72	3.56	2.88	3.76
		10	0.212	30.29	0.19	10	0.056	30.02	3.77	0.069	0.272	1.0	10.3	11	3.106	3.38	4.64	18	2.878	3.65	5.01	3.50	4.81
		10	0.212	30.29	0.19	10	0.056	30.02	3.77	0.069	0.272	1.0	10.3	11	3.106	3.38	4.64	18	2.878	3.65	5.01	3.50	4.81
		30	0.161	30.30	0.08	30	0.055	30.01	2.90	0.031	0.291	1.0	35.8	36	2.719	8.39	11.83	58	2.663	8.56	12.08	7.60	10.72
		6	0.123	30.24	0.16	6	0.080	29.99	1.53	0.060	0.251	1.0	8.6	9	3.250	3.85	5.13	10	3.169	3.94	5.26	4.17	5.56
40	95I40004	6	0.092	40.50	0.12	6	0.037	40.55	2.47	0.041	-0.051	1.0	6.6	7	3.499	6.70	7.05	10	3.169	7.39	7.79	7.81	8.23
		10	0.081	40.51	0.07	10	0.041	40.53	1.97	0.029	-0.022	1.0	13.4	14	2.977	11.43	11.69	18	2.878	11.82	12.09	11.34	11.60
		10	0.081	40.51	0.07	10	0.041	40.53	1.97	0.029	-0.022	1.0	13.4	14	2.977	11.43	11.69	18	2.878	11.82	12.09	11.34	11.60
		30	0.091	40.52	0.04	30	0.061	40.49	1.49	0.020	0.032	1.0	50.7	51	2.676	18.10	18.69	58	2.663	18.18	18.78	16.14	16.67
		6	0.125	40.51	0.16	6	0.068	40.44	1.84	0.058	0.073	1.0	7.7	8	3.355	4.76	5.13	10	3.169	5.04	5.43	5.32	5.74
50	95I50005	6	0.068	49.90	0.09	6	0.044	49.91	1.56	0.033	-0.007	1.0	8.5	9	3.250	9.26	9.33	10	3.169	9.50	9.56	10.03	10.10
		10	0.070	49.92	0.06	10	0.040	49.91	1.74	0.026	0.008	1.0	14.4	15	2.947	13.19	13.29	18	2.878	13.50	13.61	12.95	13.06
		10	0.070	49.92	0.06	10	0.040	49.91	1.74	0.026	0.008	1.0	14.4	15	2.947	13.19	13.29	18	2.878	13.50	13.61	12.95	13.06
		30	0.060	49.94	0.03	30	0.041	49.90	1.46	0.013	0.039	1.0	51.4	52	2.674	27.05	28.14	58	2.663	27.16	28.25	24.11	25.08
		6	0.050	49.96	0.06	6	0.021	49.88	2.36	0.022	0.081	1.0	6.7	7	3.499	11.85	12.89	10	3.169	13.08	14.23	13.82	15.04
60	95I60001	6	0.089	60.44	0.12	6	0.007	60.77	11.90	0.036	-0.325	0.5	5.1	6	3.707	1.29	3.70	10	3.169	1.51	4.33	1.60	4.57
		10	0.080	60.43	0.07	10	0.012	60.76	6.78	0.026	-0.335	0.5	9.4	10	3.169	2.04	6.17	18	2.878	2.25	6.79	2.16	6.52
		10	0.080	60.43	0.07	10	0.012	60.76	6.78	0.026	-0.335	0.5	9.4	10	3.169	2.04	6.17	18	2.878	2.25	6.79	2.16	6.52
		30	0.076	60.40	0.04	30	0.016	60.76	4.75	0.014	-0.356	0.5	31.6	32	2.738	3.71	12.88	58	2.663	3.82	13.24	3.39	11.75
		6	0.101	60.39	0.13	6	0.019	60.77	5.37	0.042	-0.377	0.5	5.3	6	3.707	0.79	3.22	10	3.169	0.93	3.76	0.98	3.97
		Parent Deadweight System				*Dependent 600 System*																	
30	95I30005	6	0.052	30.00	0.07	5	0.152	30.26	0.34	0.071	-0.264	1.0	4.8	5	4.032	2.56	3.48	9	3.250	3.18	4.32	3.44	4.68
		10	0.056	30.02	0.05	5	0.152	30.26	0.37	0.070	-0.242	1.0	4.6	5	4.032	2.68	3.53	13	3.012	3.58	4.72	3.60	4.74
		10	0.056	30.02	0.05	10	0.145	30.21	0.39	0.049	-0.192	1.0	11.7	12	3.055	5.38	6.66	18	2.878	5.71	7.06	5.48	6.78
		30	0.055	30.01	0.03	30	0.170	30.29	0.33	0.033	-0.281	1.0	35.1	36	2.719	8.10	11.26	58	2.663	8.27	11.50	7.34	10.21
		6	0.080	29.99	0.11	5	0.239	30.32	0.34	0.112	-0.331	1.0	4.8	5	4.032	1.48	2.22	9	3.250	1.84	2.75	2.00	2.98
40	95I40004	6	0.037	40.55	0.05	5	0.114	40.34	0.33	0.053	0.211	1.0	4.7	5	4.032	3.68	4.66	9	3.250	4.57	5.78	4.95	6.27
		10	0.041	40.53	0.04	5	0.114	40.34	0.36	0.053	0.192	1.0	4.5	5	4.032	3.81	4.71	13	3.012	5.10	6.31	5.12	6.34
		10	0.041	40.53	0.04	10	0.116	40.33	0.35	0.039	0.202	1.0	11.2	12	3.055	6.71	8.41	18	2.878	7.12	8.93	6.83	8.56
		30	0.061	40.49	0.03	30	0.141	40.35	0.43	0.028	0.138	1.0	39.5	40	2.704	11.36	13.19	58	2.663	11.54	13.39	10.24	11.89
		6	0.068	40.44	0.09	5	0.179	40.32	0.38	0.085	0.117	1.0	5.0	5	4.032	2.58	2.93	9	3.250	3.21	3.63	3.47	3.93
50	95I50005	6	0.044	49.91	0.06	5	0.179	49.52	0.24	0.082	0.387	1.0	4.4	5	4.032	1.85	3.02	9	3.250	2.30	3.75	2.49	4.06
		10	0.040	49.91	0.04	5	0.179	49.52	0.22	0.081	0.392	1.0	4.2	5	4.032	1.86	3.06	13	3.012	2.49	4.10	2.50	4.11
		10	0.040	49.91	0.04	10	0.149	49.57	0.27	0.049	0.342	1.0	10.3	11	3.106	4.34	6.60	18	2.878	4.69	7.12	4.50	6.83
		30	0.041	49.90	0.02	30	0.120	49.59	0.34	0.023	0.311	1.0	35.7	36	2.719	10.93	15.87	58	2.663	11.16	16.21	9.91	14.39
		6	0.021	49.88	0.03	5	0.055	49.56	0.38	0.026	0.319	1.0	5.0	5	4.032	6.48	9.51	9	3.250	8.04	11.80	8.71	12.79
60	95I60001	6	0.007	60.77	0.01	5	0.167	60.06	0.04	0.075	0.705	0.5	4.0	5	4.032	-0.68	1.66	9	3.250	-0.85	2.06	-0.92	2.23
		10	0.012	60.76	0.01	5	0.167	60.06	0.07	0.075	0.705	0.5	4.0	5	4.032	-0.68	1.66	13	3.012	-0.91	2.22	-0.91	2.23
		10	0.012	60.76	0.01	10	0.185	60.09	0.06	0.059	0.675	0.5	9.1	10	3.169	-0.94	2.69	18	2.878	-1.03	2.96	-0.99	2.84
		30	0.016	60.76	0.01	30	0.127	60.09	0.13	0.023	0.666	0.5	29.9	30	2.750	-2.58	7.78	58	2.663	-2.66	8.03	-2.36	7.13
		6	0.019	60.77	0.02	5	0.045	60.12	0.42	0.022	0.647	0.5	5.2	6	3.707	-1.84	6.26	9	3.250	-2.10	7.14	-2.27	7.74

Determining the right strategy for improving Cc depends on which factor provides the greatest leverage. The following summary Tables 5.4 and 5.5 identify candidate causes of non-capability and their proposed candidate solutions. The causes are divided between mean offset causes and variation causes. These causes are not rank ordered based on leverage, but rather provide suggestion of focus areas for directing improvement attention.

Table 5.4: Non-Calibration Capability due to Mean Offset (*Condition 1*)

Cause for Mean Offset	Proposed Solution
Excessive block variation	Use blocks of higher block uniformity (Large grade blocks)
Standard drifting over time	• Increasing n_x used to calibrate internal blocks used for initial calibration of production testers prior to shipment • Cross-reference standards testers with more stable, lower variation deadweight tester
Sampling error in initial tester calibration	Increasing n_y of the production testers for initial calibration to the reference standard
Adjustment error	Increase number of measurements by service men when adjusting the calibration settings on customer units in the field
• Non-controlled state of tester • Mechanical wear over time	Use statistical control chart methods for tester to detect mean shifts with appropriate cause correction
Measurement Round-Off error	Increase the significant digits of the read-out display in parent and dependent system
True Mean Offset	Dependent system mean adjustment by 'dipping' with caution of over-control

Table 5.5: Non-Calibration Capability due to Variation (*Condition 2*)

Cause for Variation	Proposed Solution
Excessive block variation	Use higher uniformity blocks (Large grade)
Low number of subgroup measurements	Increase n for system with largest variance
• Non-controlled state of tester • Mechanical wear over time	Use statistical control chart methods for tester to detect patterns in s or R with appropriate cause correction
Mechanism Error	• SPC during tester production • Enhanced tester technology

Chapter 6 Metrics for Process Capability

A capability assessment of a process, such as the one for manufacturing hardness test blocks, aims to determine if the random variability in the process is sufficiently small such that the product can be consistently manufactured to meet customer needs. The customer needs are 'translated' by the manufacturer in the form of product specifications, for example maximum and minimum hardness values or allowable hardness range (Refer to Table 1.1).

Two capability metrics are proposed for the manufacture of hardness test blocks:

- the process capability, C_{pk}
- the range capability, C_R

The C_{pk} will seem familiar to the student of statistical process control methods. The C_{pk} is discussed with regard to its application to Instron's block manufacturing process. The range capability, C_R is derived by the author to satisfy the context by which the customer is currently accustomed to assessing hardness measurement variation: the allowable range per block [4].

6.1 The Process Capability Index, Cpk

The Cpk index measures if the average block hardness is within the acceptable hardness specification limits desired for that process. In addition, the index detects if the average block hardness for the process are centered (or otherwise shifted off-center) between the upper and lower hardness specification limits (USL and LSL). The Cpk index thus detects mean shifts in the process.

The Cpk is given as [14]:

$$C_{pk} = \min\left\{\frac{USL-\overline{X}}{3\cdot\frac{s}{n^{1/2}}},\frac{\overline{X}-LSL}{3\cdot\frac{s}{n^{1/2}}}\right\} \tag{33}$$

The reader will note that a normal probability distribution for the block average is assumed based on the application of the Central Limit Theorem. Refer to Chapter 3.0. Hence, a z-statistic of 3 is used as the critical value of the normal test statistic, representing 99.7% confidence with perfect estimates of s and $\overline{X}$.

Current benchmarks for an acceptable Cpk are 1.33 [18]. Rockwell International boasts a plant with a minimum C_{pk} of 4.09 [28]. In order to account for sampling variation in Cpk, critical values greater than 1 are desired.

While the goal of the metric is to measure the process performance on achieving the target hardness of the block, the variation of the tester used for the hardness measurement plays a significant role in the Cpk value. It was demonstrated in Chapter 2 that for Large blocks the tester accounted for the major portion of the measurement standard deviation, s. Hence, the metric emphasizes the importance of tester control used for process measurements.

6.2 Cpk achieved through Feedback Control in Heat Treating

From discussions with Instron management, the ability to achieve an acceptable average hardness is not deemed a significant challenge since the specification limits are quite large. In general the specification limits to the customer are *2 Rockwell points* on either side of the nominal hardness. The reason why this is not held as very critical is that if a process batch of blocks falls outside the specification limit, it can be substituted into another nominal hardness level or part number. However, this salvage policy does not account for the considerable costs brought about by increased inventory levels and expediting measures in order to satisfy customer demand.

The ultimate tactical burden for achieving Cpk lies with the heat treater. Fig. 6-1 shows the feedback control representation of the tempering process used for achieving target hardness in heat treating. The hardening and tempering of blocks is a batch process of 25 to 100 blocks per batch depending on block type. When the blocks are initially placed in the tempering furnace they are at their maximum hardness, resulting from austenizing and quenching. The hardness of the blocks is lowered during tempering. The process batch is then tempered for a fixed time interval of 2 hours at a specific temperature. This temperature parameter is determined from the heat treaters experience or is estimated from a time-temperature curve to yield a specific hardness.

For the general classification of tool steels, of which the specialty steels of hardness blocks are a subset, hardness during tempering is a function of both *time and temperature* [20][11]. Time only becomes a significant parameter for hardness levels greater than HRC 45 with tempering temperatures greater than 450 C. Refer to Appendix H, Figure H-1, for a typical tempering curve, that relates hardness to both time and temperature parameters; the upper temperature curve is deemed sufficient as a practical heat treating reference.

After the two hour interval, the heat treater removes a sample from the batch which is air-cooled and prepared for hardness tests on the surface. If these in-process sample measurements indicate that the batch is within the target hardness range ($X_{min} < \bar{X}1 < X_{max}$), the batch is removed from the furnace and rapidly cooled. A final post-process hardness measurement ($\bar{X}2$) is conducted by the heat treater on a different batch sample prior to shipment to Instron.

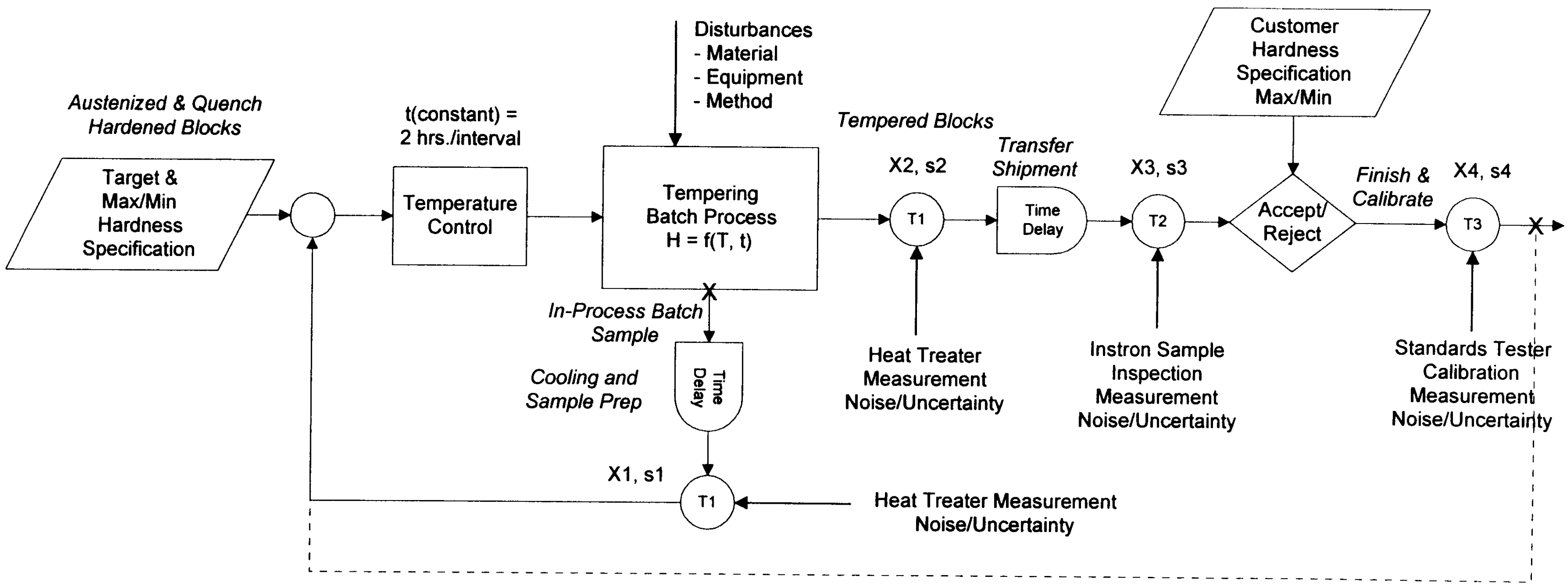
A final acceptance sample measurement is conducted by Instron upon receipt of the batch ($\bar{X}3$). The ultimate determination of block hardness occurs after the blocks have been finished and are calibrated in the Instron Standards Laboratory ($\bar{X}4$).

[11] Thelning [20] introduces empirical equations of the form:

$$H = T(k + \log t)$$

where, T = Temperature in Kelvin; t = Time in hours; k = constant

Figure 6-1: Feedback Control of Tempering for Target Nominal Hardness and Process Capability



The heat treater selects the initial temperature parameter for the tempering process target based on the upper end of the allowable hardness specification, as a safety margin rule. On average, two to four 2-hour intervals of adjusted (decreasing) temperature parameters are required in order to achieve the target sample hardness [interviews]. The heat treater tries to aim for the mid-point of the hardness spec range [interviews].

As of December 1995, the heat treater was given the same allowable hardness range for the nominal target hardness for the tempering process as was used in the Instron Standards laboratory for calibration sorting.

6.2.1 Challenges to Achieving C_{pk}

Several factors make achieving the Cpk more complex despite the large specification range:

- Block hardness measurements for batch qualification are conducted by three sets of different testers: the heat treaters testers (T1), the Instron sample inspection (T2) and the block calibration (T3). In order to avoid measurement discrepancies, all testers must be calibrated with respect to the final Standards tester (X4). Calibration errors are natural, given the variation demonstrated earlier.
- Surface condition of the measured samples is not the same for each measurement sampling at each inspection stage. The heat treatment scale must be removed by the heat treater. Differences in flatness and parallelism also account for measurement noise.
- If the heat treater is under time pressure, the tendency by the safety margin rule is to err on the high side of the specification tolerance. The remedy for non-conformance is further tempering, as opposed to the potential penalty of batch rejection.
- Heat treater decisions are governed by comparing *individual measurements* of n=5 to the specification limits, instead of the sample average. Individual measurements are expected to have large variation due to poor surface condition of the tempering samples.
- For HRC 50 and above: The process batch continues to soften in the furnace while the in-process sample is being measured. The time delay associated with cooling and preparing the block sample for measurement can reach 30 minutes. During this time the batch is still in the tempering furnace. The tempering curve of Appendix H shows that this time delay is not insignificant at high temperatures. A 0.5 Rockwell point discrepancy can be expected for tempering temperatures greter than 1000 F. Refer to Figure H-1. The heat treater must therefore account for the difference in hardness between the lot sample (coupon) and the batch.

6.2.2 Proposed C_{pk} Enablers

Several modifications to the existing methods are proposed in order to improve Cpk:

- Develop tempering curves from historical data points for the particular material employed for test blocks. This will allow the reduction of required tempering intervals, as the initial estimate of the output hardness will be improved.
- Ensure that the heat treater and in-process testers are of adequate state of technology and are well maintained so as to minimize the variation attributed to the tester. As a result, the reduction in measurement variation (denominator) will serve to enhance capability.

- All in-process testers should be frequently monitored in accordance with the standardizing testers used for block calibration. This aim is to avoid drifting of the relative calibration settings. This cross-referencing can be achieved with sets of master reference blocks calibrated and dispatched by the Instron Standards lab.
- Develop consistent methods for surface preparation both at the heat treater and Instron sample inspection. The consistency in surface condition, albeit poor, allows for the determination (over time) if the surface condition accounts for mean shifts or added variability.
- Increase the number of measurements at all stages from n=5 to n=10. This strategy can be used to compensate for the enhanced variation due to poor surface quality of in-process samples and tester variation in order to improve the estimates of the block average.
- Use averages not individual measurements for assessing the target hardness. The heat treater and the production associate performing the in-house sample inspections must be cautious in reacting to outliers in a set of 5 measurements. The outliers are natural given the poor surface quality or otherwise. The average allows extremes to cancel each other out. All decisions with respect to the acceptable hardness specification should be based on the sample average.
- MEASURE, record and track the measurement averages $\overline{X}$ and standard deviations, s, at each measurement station for each batch over time. A simple run chart is a great process control tool for relating different sets of testers. If the heat treater finds consistent discrepancies between $\overline{X}1$ and $\overline{X}2$, investigation into the root causes must be aimed at both disturbances acting on the tempering process, as well as noise disturbances acting on the tester.
- Use offsetting spec ranges to compensate for in-process measurement error. In effect, the heat treater would be given tighter specifications than those of the Standards laboratory which reflect the customer requirements. This tighter working hardness range may increase the number of tempering intervals due to caution and subsequently cost. However, the improved control measures proposed thusfar are expected to reduce the number of required tempering intervals.

It is shown in this example, that both the production systems, e.g. the tempering process, and measurement systems must be closely linked in order to achieve a satisfactory Cpk.

6.3 A simple Cpk simulation from actual data

A simulation was conducted using available data which we assume to represent 'typical' values of combined tester and block variation. C_{pk} values are calculated at different measurement quantities, n = 5 and n = 10.

In addition, using the criteria of an allowable $C_{pk} = 1.5$, the minimum allowable difference between a specification limit and average reading, SL-$\overline{X}$, is determined at n = 5 and n = 10. Refer to Table 6.1.

The simulation shows that on occasion the standard deviation at n = 10 is larger than at n = 5. The reader should not be tempted to pursue whatever sample size gives a smaller standard deviation and hence higher capability. Rather, an increased number of measurements gives an improved estimate of the actual block standard deviation.

The simple simulation shows that most Cpk values are considerably larger than 1, except for one instance of HRC 25 block measurements.

In general, the standard deviations for the samples show that to achieve a reliable Cpk of 1.5, the minimum allowable difference USL-$\overline{X}$ or $\overline{X}$-LSL may not be less than 0.5 for any hardness level. This is concluded on average based on the assumption that the simulation values are 'typical' of the process.

6.4 Using Calibration Capability to achieve Process Capability

From the above simulation we may be tempted to conclude that as long as the heat treater does not produce a block within 0.5 of either upper or lower customer specification limit, the customer spec limits should not be violated in the vast majority of time. Let us take the case when customer spec limits are +/- 2.0 points above and below the nominal hardness value e.g . HRC 47 to 43 for a nominal HRC 45. If the heat treater supplies a specification range of 46.5 to 43.5, the process capability with respect to the customer requirements is guaranteed. Right ? Well, not exactly. If there exists a calibration offset between the Standards testers and the heat treater tester, there is still a potential for a violation of the customer specification limits. Therefore, the policies and cross-referencing methods for aligning the mean hardness levels of the Standards and heat treater testers are critical.

On the basis that a system is implemented by which these testers are well maintained and cross-referenced, we may be able to bound the likely mean difference ($\overline{X}$4-$\overline{X}$1) between testers using the *calibration capability*. This will require monitoring and recording the cross-referenced $\overline{X}$ and s behavior of both Standards and heat treater testers over time. Refer to Figure 6-2.

Calibration Capability between the Standards tester and heat treater tester may be successfully determined to be greater than 1 using a high-quality grade test block. Clearly, the mean offset $\overline{X}$4-$\overline{X}$1 may not be as large as the allowable standard half-tolerances given on the reference calibration blocks e.g $\delta x = 1.0$ at HRC 45. However, a Cc of 1 means that the mean difference ($\overline{X}$4-$\overline{X}$1) is consistently less than $\delta x = 1.0$ [See Equation (31)]. For such a half-tolerance, 1.5 points of the customer specification limit would be absorbed leaving an unacceptable window of +/- 0.50 HRC for the heat treater to aim for in tempering.

However, suppose that the Cc was determined to be consistently greater than 2.0 for 3 separate trials of Cc at n=10 measurements, as we have seen in several simulations ? In this case, we could bound the maximum level of the mean offset $\overline{X}$4-$\overline{X}$1 as 0.5 Rockwell points, assuming good estimates of the standard deviations and averages. In this case, the heat treater could be given specification limits that lie within 0.5 + 0.5 = 1 Rockwell points inside the customer specification limits. As a result, the heat treater would target HRC 46 to 44 leaving a 2.0 point window for tempering to hardness.

A 2.0 point window is within the limits of feasibility for the heat treater. The heat treater's incentive for working with a tighter specification window is the assurance that batch rejection is extremely unlikely, assuming that everyone in the process chain ensures that their testers are in control.

This hardness control scenario demonstrated how the process capability and calibration capability measures can be used to make policy decisions in order to enhance the working reliability of the process and in turn the quality of the reference standard supplied to the customer.

Figure 6-2: Relationship of Tempering Target Window to Customer Specification Limits

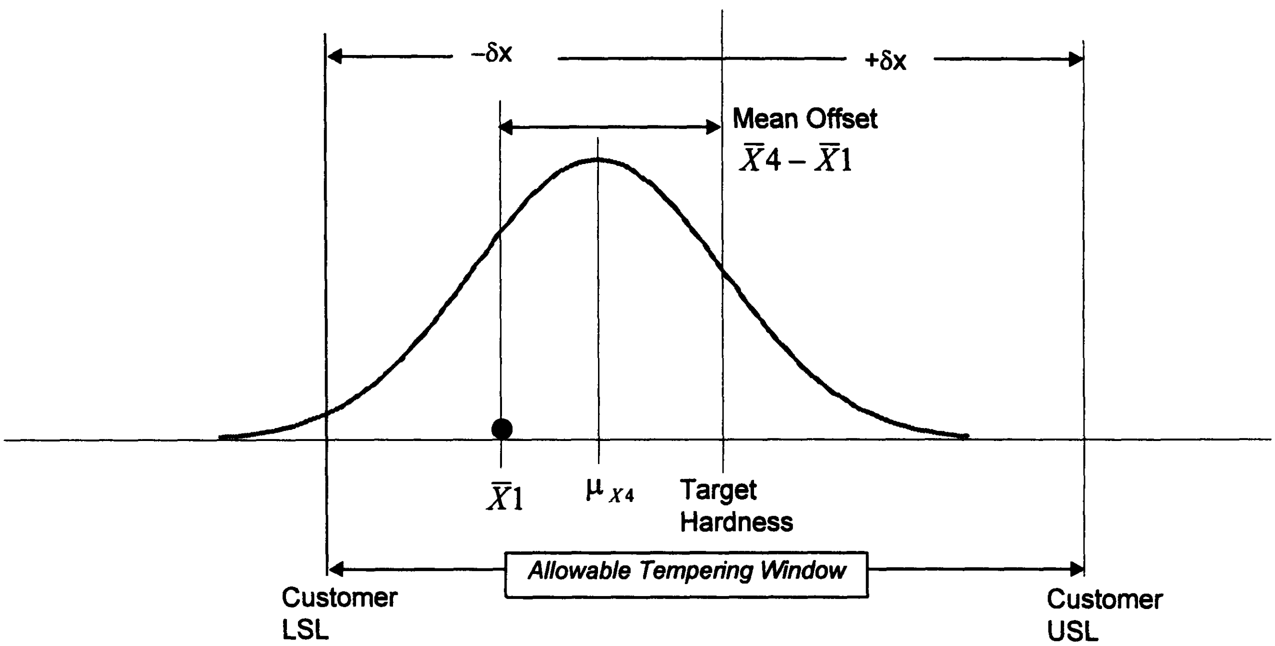


Table 6.1: Simulation of Process Capability and Range Capability

HRC Nominal	n	Average $\overline{X}4$	Std. Dev. s4	s4/sqrt(n)	USL	LSL	min. SL-$\overline{X}$	**Cpk**	min (SL-$\overline{X}$) for Cpk = 1.5	Rspec	C_R
Regular Blocks											
25	5	26.12	0.164	0.073	27.0	23.0	0.9	**4.0**	0.33	1.0	**1.5**
	10	26.34	0.313	0.099	27.0	23.0	0.7	**2.2**	0.45	1.0	**0.8**
	5	26.56	0.313	0.140	27.0	23.0	0.4	**1.0**	0.63	1.0	**0.8**
45	5	45.66	0.152	0.068	47.0	43.0	1.3	**6.6**	0.31	1.0	**1.6**
	10	45.81	0.197	0.062	47.0	43.0	1.2	**6.4**	0.28	1.0	**1.3**
	5	45.92	0.192	0.086	47.0	43.0	1.1	**4.2**	0.39	1.0	**1.3**
63	5	63.48	0.192	0.086	65.0	61.0	1.5	**5.9**	0.39	0.5	**0.7**
	10	63.50	0.133	0.042	65.0	61.0	1.5	**11.9**	0.19	0.5	**0.9**
	5	63.50	0.122	0.055	65.0	61.0	1.5	**9.2**	0.25	0.5	**1.0**
Large Blocks											
30	5	30.26	0.152	0.068	32.0	28.0	1.7	**8.5**	0.31	1.0	**1.6**
	10	30.21	0.145	0.046	32.0	28.0	1.8	**13.0**	0.21	1.0	**1.7**
	5	30.32	0.239	0.107	32.0	28.0	1.7	**5.2**	0.48	1.0	**1.0**
40	5	40.34	0.114	0.051	42.0	38.0	1.7	**10.9**	0.23	1.0	**2.2**
	10	40.33	0.116	0.037	42.0	38.0	1.7	**15.2**	0.17	1.0	**2.2**
	5	40.32	0.179	0.080	42.0	38.0	1.7	**7.0**	0.36	1.0	**1.4**
50	5	49.52	0.179	0.080	52.0	48.0	1.5	**6.3**	0.36	1.0	**1.4**
	10	49.57	0.149	0.047	52.0	48.0	1.6	**11.1**	0.21	1.0	**1.7**
	5	49.56	0.055	0.025	52.0	48.0	1.6	**21.1**	0.11	1.0	**4.5**
60	5	60.06	0.167	0.075	62.0	58.0	1.9	**8.7**	0.34	0.5	**0.7**
	10	60.09	0.185	0.059	62.0	58.0	1.9	**10.9**	0.26	0.5	**0.7**
	5	60.12	0.045	0.020	62.0	58.0	1.9	**31.1**	0.09	0.5	**2.8**

Chapter 7 The Range Capability Index, C_R

Manufacturers and customers of test blocks alike currently evaluate the measurement variation of the block on the basis of the allowable measurement ranges defined by the applicable standard specification e.g. ASTM E-18 per Table 1.1 [4] is thus appropriate to evaluate the performance of the block manufacturing process with regard to this quality characteristic.

The Range Capability Index, C_R, is proposed as the appropriate metric with regard to standard range specifications:

$$C_R = \frac{Rspec}{4\sigma} \geq 1 \tag{34}$$

The author developed this metric on the basis of the classic process capability index, Cp which is written [14] as,

$$Cp = \frac{USL - LSL}{6\sigma} \geq 1 \tag{35}$$

The reader will note that in equation (35) the difference between upper specification limit and lower specification limit *USL-LSL* is equivalent to a specification on *Range*. Both C_R and Cp metrics work on the basis that the subject measurement distribution is centered between the specification limits. The generally false centering assumption represents a serious deficiency of the Cp index, prompting the more informative Cpk index. However, in the C_R assessment if the total measurement variation fits inside the specification range (the difference between upper and lower specification limits), the centering deficiency is not of consequence.

The C_R index determines if the majority of the normal distribution that describes the individual and potential hardness measurements fits inside the range window. Refer to Fig. 7-1 below. Normality of individual measurements was confirmed to be a reasonable approximation in Chapter 3.

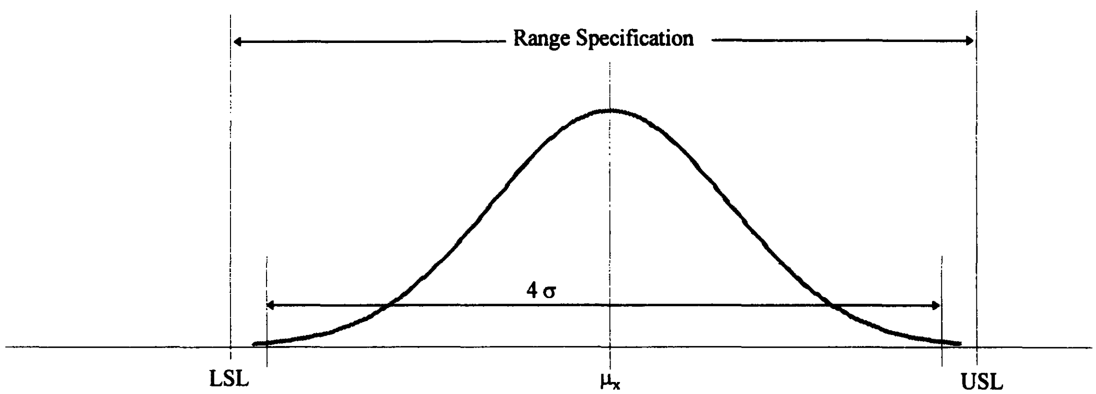
C_R is calculated on the basis of 95.5 % probability (4-sigma, two-tail) vs. 99.73% (6-sigma, two-tail) for Cp that at all the measurements on a block will fall inside of the specification range. The 95% level is common for reporting methods of uncertainty in measuring equipment [1,9].

For a $C_R = 1$, there exists a 95% probability that the hardness measurements will lie inside the range specification, assuming that:

- the sample standard deviation detected for the process, s, is a good estimate for the true block standard deviation, σ.
- the sample standard deviation is 'typical' or 'average' for the process.
- the measurement variation is normally distributed. Refer to Chapter 3.

For small sample sizes less than 30, there exists larger sampling error. The first two assumptions can be achieved if the s-values are taken as the average standard deviation $\bar{s}$ taken from the appropriate control chart for standard deviation s. The average std. deviation $\bar{s}$ of numerous (N>30) blocks of small measurement subgroups of n = 5 or 6 is a strong estimate of the true process σ. [Refer to Chapter 8].

Figure 7-1: Range Capability for 4-sigma fit inside the Range Specification



If it is known that the σ-estimates and thus C_R are to sampling error due to small sample sizes, it is typical to increase the critical C_R value, for example $C_R > 1.33$ [18].

It should be noted that the standard deviation, s, in practicality represents the total measurement variation of a local measurement system. The block variation is masked by measurement noise. Thus, for different measurement variation of the parent and dependent tester systems respectively, it is possible that the customer will measure a different (larger) C_R for an identical block previously characterized by the manufacturer (Instron), even though it was derived from the same process lot.

7.1 A C_R simulation from actual data

A simulation for C_R is shown in Table 6.1 on the assumption that the sample blocks measurements yield standard deviations that are "typical" and good estimates of σ.

The simulation demonstrates that the base criteria for C_R=1 may be not be currently fulfilled with state of total measurement variation using a 500 tester and Regular blocks at HRC 25, 45 and 63. There is an indication that the Large blocks have considerably improved the Range capability, despite some potential difficulty at HRC60.

In general, the purpose of the Range Capability, like the other capabiltiy indices, is to be used as a tool for measuring and driving improvement rather than confirming compliance. For the 4-

sigma range coverage (C_R=1), there still exists a 5% probability that a measurement set can fall outside of the range specification.

7.2 An Exercise for the Critical C_R Value

The understanding of the C_R index is supported by the following exercise based on empirical data.

From the measured R/s behavior of a 600 tester as depicted in Fig. 4-6, it can be concluded that for small measurement sample sizes of 5 or 6, that the maximum theoretical (95%) R/s ratio is 4.

Hence it follows:

Since, $R/s < 4$ for $n \leq 6$ (See Fig. 4-6)

Therefore, $s < R/4$ or $s_{max} = R_{max}/4$

Substituting into the CR expression (34),

$$C_R = R_{SPEC} / (4 * s_{max}) \geq 1$$

$$C_R = R_{SPEC} / (4 * (R_{max}/4) \geq 1$$

$$C_R = R_{SPEC} / R_{max} \geq 1$$

Therefore, a C_R of 1 means that the range specification just equals the maximum sample range.

Part III Statistical Methods for Process Control

Chapter 8 Statistical Process Control for Manufacture of Hardness Test Blocks

Thusfar, we have focused on measures of calibration capability and process capability. Process capability is defined, in short, as the ability of the process to reliably produce parts that conform to engineering specifications. The assumption is made that these specifications properly reflect customer needs.

Determination of capability requires assurance that the process is in a state of control: '... it is both inappropriate and statistically invalid to assess process capability with respect to conformance to specifications without being reasonably assured of having good statistical control. Although control certainly does not imply conformance (to specifications), it is a necessary prerequisite to the proper assessment of conformance" [14]. The requirement of statistical control before capability, while strongly advocated by this author, is still subject to debate in the SPC community [30].

The capability metrics contain representative quantification of the system's output performance taken at an instant of time: the measurements averages, $\overline{X}$, and the sample standard deviations, s. The capability metrics are only useful if they represent reliable predictions of future performance. Unless the $\overline{X}$ and s statistics are 'typical' of the system under study and the system is not known to vary or drift significantly over time, the capability metrics may be misleading about the state of the process variation.

In this chapter the feasibility and utility of statistical methods for process control (termed SPC) for the manufacture of hardness test blocks are investigated. These statistical process control methods are tools for determining if the process is operating in a state of statistical control and to gage the improvement of the process over time.

Four main challenges are addressed in controlling block hardness characteristics:

- Measurement noise from the hardness testers in detecting out-of-control conditions of block process
- Small batch sizes for each part number/hardness level (*short runs*)
- Minimizing required control chart quantity and charting effort, as a result of the variation behavior particular to hardness level/part number (Refer to Chapter 2)
- Distinguishing systemic, special causes for variation of the block process from those of the measurement testers

An SPC strategy that addresses these constraints is proposed.

8.1 Control of Measurement Testers for Statistical Control of Block Process

The author extends the requirement of statistical control to the measurement systems used to test the blocks in production, the Rockwell testers of the Standards Laboratory. "The measuring system may also include poorly controlled processes that lead to (random) variations in the system output"[1].

Figure 8-1 depicts a *process control perspective* of the Instron test block manufacturing and calibration process [19]. The standards testers are an integral part of the producing the final customer product: a reference material + reference information ($\overline{X}$, s, R). In addition, it was determined in Chapter 2 that the tester system is a major source of measurement variation in the Rockwell measurement of test blocks. Because they are an integral part of the total process and they account for a large portion of the measurement variation, the testers must also be ascertained to be in statistical control. "Understanding and quantifying this measurement error is an important aspect that is often overlooked when one is charting the performance of a process" [18].

The testers in the Instron Standards laboratory currently play two important roles. First they calibrate the test blocks for the customer providing hardness reference information. Secondly, they serve as a diagnostic instrument for the manufacturing process to determine that the process satisfies customer-based specifications. It will be explored if they can perform a third role: to serve as a diagnostic instrument in order to achieve statistical process control of the total block manufacturing process.

8.2 Definitions of Statistical Control: Random and Systemic Variation

Variation of a manufacturing process can be divided into two types: random and systemic. Systemic variation is also referred to as special-cause variation (Deming), or disturbances, meaning that its sources can be traced to known causes that disturb the natural random behavior of the system. The random variation, also called common-cause or natural variation, distinguishes the variation that can be expected for the system and behaves in a random fashion between bounded limits. The variation attributed to the tester system used to measure the test blocks is commonly referred to as measurement noise. Measurement noise also consists of random and systemic variation.

A process or measurement system is in statistical control when non-random systemic disturbances are not present.

The control chart, the diagnostic tool of SPC, has the basic purpose "to provide evidence of whether a process has been operating in a state of statistical control and to signal the presence of special causes of variation so that corrective action can be taken" [18].

Random variation may be modeled by probability distributions of random variables, such as the Gaussian normal, by which the block measurement variation is characterized [See Chapter 4]. Special-cause variation does not behave like a random variable. It is this non-random characteristic that allow control charts, to detect that systemic, special-cause variation is present.

8.3 Feedback control system view of SPC applied to the test block process

The purpose of control charts is to examine the output data for the process, and to detect if disturbances are present. It is further implied that the special cause for any detected disturbance should be diagnosed and eliminated. This detection and corrective action process for achieving statistical control is depicted as a SPC feedback loop containing five steps: *Observation, Evaluation, Diagnosis, Decision, Implementation* [14]. Refer to Figure 8-1. Table 8.1 summarizes the definition and tools for each step. The data collection (observation) is conducted using the testers in the Instron Standards laboratory.

Evaluation of the feasibility of SPC for the Instron block manufacturing process must therefore extend beyond the enablers and barriers of control charting. The organizational capabilities, infrastructure and skills required for all of the *five steps of SPC feedback* leading to corrective action for the removal of process disturbances must be addressed. "The importance of cooperation between the manufacturer and the calibrating laboratory cannot sufficiently be underlined" [3].

Table 8.1: Five Steps of SPC Feedback [14, 24]

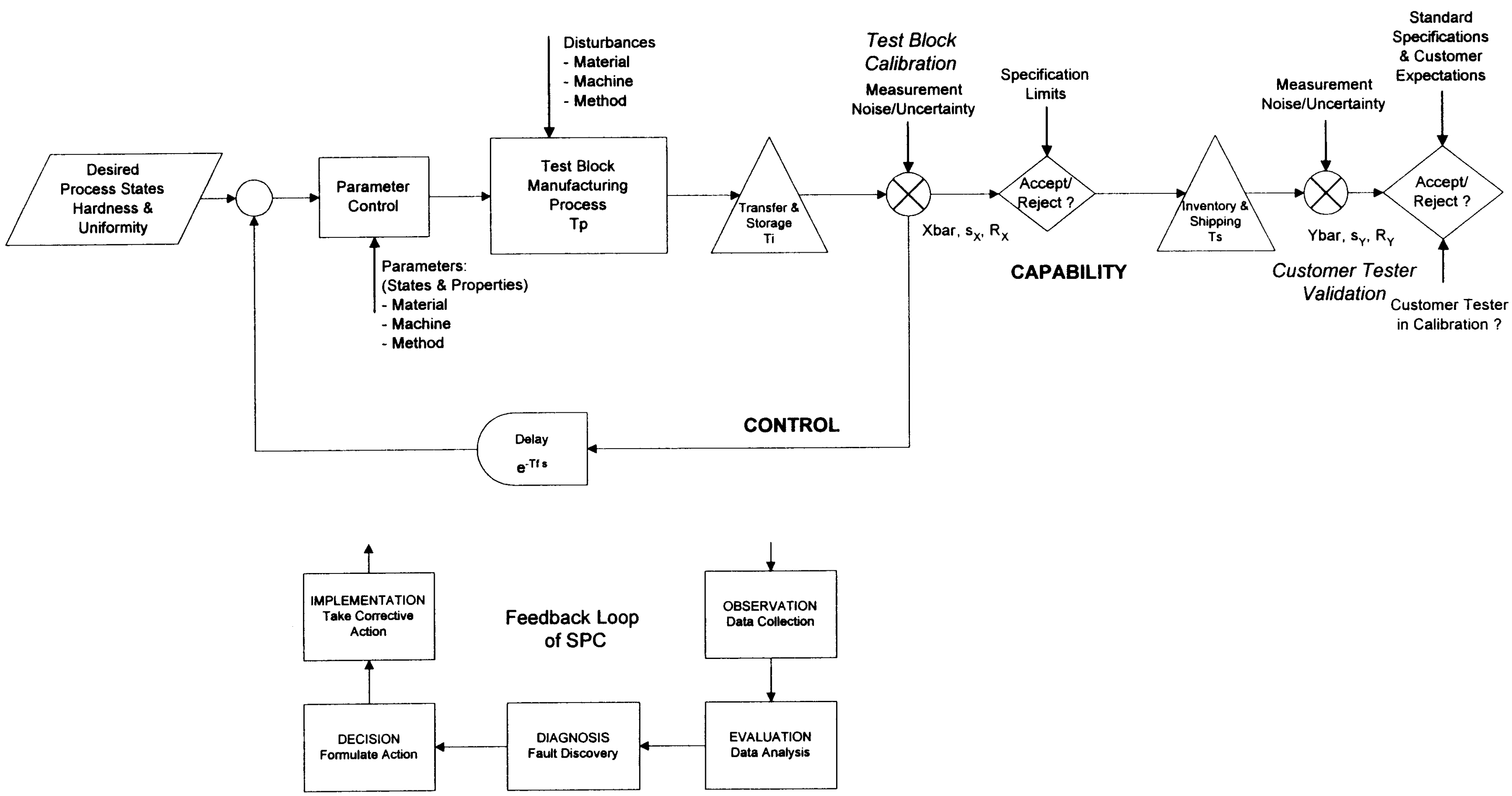
Feedback Step	Definition	Available Tools
Observation	Data collection	Segmentation Data Acquisition Systems
Evaluation	Data analysis	Control Charts Histograms Scatter Diagrams Pareto Diagrams
Diagnosis	Fault discovery	Cause-and-Effect (Ishikawa) Diagrams Traceability Records
Decision	Formulate action	Action Planning
Implementation	Take action	Confirmation Experiments Procedures Specifications

8.4 Selection of Summary Statistics for Control Charting

The next question to be addressed: 'Which characteristic statistics should be charted ?"

The author maintains the following key statistics are relevant in determining statistical control for block manufacture. In addition, they directly map into the customer requirements (Cc), existing standards specifications and capability metrics:

Figure 8-1: Process Control System View of Test Block Manufacturing



- Calibration average, $\overline{X}$, for individual blocks
- Calibration range, R, for individual blocks
- Calibration standard deviation, s , for individual blocks

The $\overline{X}$ *-chart* for calibration averages would primarily determine the how well the heat treater is able to consistently zero-in on the nominal hardness via the feedback control sub-process of tempering discussed in Chapter 6. It would also serve as a detector of other currently unknown systemic causes for shifts in the mean hardness, some potentially attributable to the tester e.g. worn-out indenter.

Statistical control of calibration averages is the alternative to sorting blocks to their nominal hardness level after heat treating. Clearly, the costs of excess inventory and the impeded delivery responsiveness of the latter policy make statistical control of block averages the preferred alternative.

The *standard deviation* is in many ways a superior metric of block variation than range for control charting. Standard deviation is less susceptible to outliers and sample size [16]. The average standard deviation, $\underline{s}$, can be directly taken from an s-chart and applied in defining other key statistical metrics on process and product performance, such as the Cc, Cpk and C_R.

However, the *range* is still the means by which the governing standard in the U.S. for Rockwell hardness testing, ASTM E-18, characterizes variation [4]. It is assumed that customers' general expectations are for conformance to the practices used by this standard. *Range* is also fundamentally easier to compute than standard deviation. Note that the Range Capability, C_R, requires a best estimate of s [See Chapter 7] .

Therefore, the author reluctantly advocates initially charting *both* standard deviation and range, each for their own merit.

Over time, if the R/s relationship becomes known within predictable limits, the range chart may be phased out [Refer to Section 4.7]. A tight R/s behavior for small sample sizes as depicted in Figure 4-6 would potentially allow the standard deviation to serve as a sufficient means for statistical variation control of the process.

8.5 Enablers to Statistical Process Control for Test Block Manufacture

8.5.1 Data collection through calibration measurements

Every block supplied by Instron is calibrated in the Standards Laboratory by n=6 measurements per block using the latest 600 tester technology. The digital 600 tester, depicted in Appendix A, automatically calculates summary statistics of mean, range and standard deviation. Currently, these summary statistics of measurement average and measurement range are manually recorded on the customer calibration certificate.

The 600 tester also features a RS232 data port that allows the measurement data in ASCII format to be downloaded to an external database. Each measurement data stream is accompanied by a descriptor for the particular Rockwell scale measured, e.g. 'HRC' for Rockwell C -scale. This descriptor is an ideal sort variable for a database to manage control chart data. These features are offered in the 600 tester to allow customers to conduct SPC from product hardness measurements.

Thus, a suitable foundation for data collection within the flow of production therefore already exists. This is particularly advantageous since Rockwell testing is a destructive test; additional measurements would otherwise use up the measurement area available to the customer.

8.5.2 Rational subgrouping

The calibration subgroups of $n = 6$ measurements for an individual block serve as ideal sample sizes for control charting. The subgroup passes the criteria established by Devor et al [14]:

- Subgroups should be subject to common-cause variation
- Subgroups should ensure the presence of normal distribution for the sample means. In practice, sample sizes of 4 or more generally ensure a good approximation to normality.
- Subgroups should ensure good sensitivity to the detection of special/assignable causes.
- Subgroups should be small enough to be economically appealing from a collection and measurement standpoint.

The central limit theorem can be directly applied on subgroup averages such that the control limits on the $\overline{X}$, R and s chart are determined on the basis of the normal probability distribution. Refer to the discussion on normality of Chapter 4.

8.5.3 Specification Limits on Control Charts

In many industrial SPC applications, conformance to specifications and statistical control each require different sets of statistical data. The former deals with populations of individual measurements for comparison to specification limits; the latter traditionally works with summary statistics for a subgroup made up of individual items, one measurement per item. As a result, Devor states that "we should never place tolerance/specification limits on an $\overline{X}$-chart" [14]. However, this not the case for the sampling scheme proposed herein for test blocks.

The summary statistics of $\overline{X}$, R and s are representative of the individual block, as are the specifications. Hence, the author maintains that in this case it is acceptable and potentially useful to place specification limits on $\overline{X}$, R and s on the control chart, as long as the users do not confuse the identity and meaning of control limits vs. specification limits. There is the potential to directly witness the how capability and control interact as a means of learning.

8.5.4 Responsiveness of Shewhart Control Charts

Because the subgroups of $n = 6$ each represent individual blocks, the responsiveness in detection of special causes variation is ensured using the simple Shewhart $\overline{X}$, R and S control charts. Sensitivity to changes in the process are also gained by charting every block calibrated, as opposed to occasional samples.

Often more elaborate and complex charting methods are necessary in manufacturing environments of small production quantities (short runs) to attain chart sensitivity, for example CUSUM or EWMA charting techniques [14]. As a result, often 'short run SPC is advanced SPC' [31].

The common eight tests for detecting out-of-control conditions from patterns on Shewhart control charts can thus be applied [14].

Refer to Figure 8-2 for an example of the Shewhart control charts based on subgroups of n=6 calibration measurements per data point.

8.6 Barriers to Statistical Process Control for Test Block Manufacture

8.6.1 Independent charts required per hardness level

It was determined in Chapter 2 that the measurement variation response for the Instron 500 and 600 testers is a function of hardness level being measured (or the displacement of penetration). Refer to Figure 2-6. A robust model for conversion of statistics between hardnesses is currently not available. Therefore, the control charts for each hardness level must be treated independently.

8.6.2 Many part numbers for many hardness scales and levels = many charts

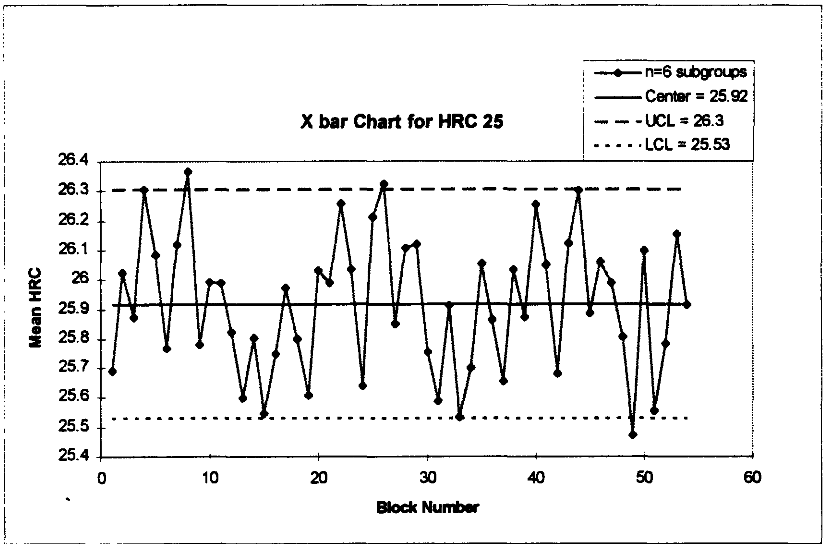
Each combination of Rockwell scale and hardness level contained in it has a different part number. Calibration is performed per part number using the exact scale designated for customer validation. Conversions between superficial and regular scales are regarded as approximate only. Therefore, since variability is a function of the tester response, it follows that each part number requires its own chart set. That would not be a problem if the set of possible part numbers were small.

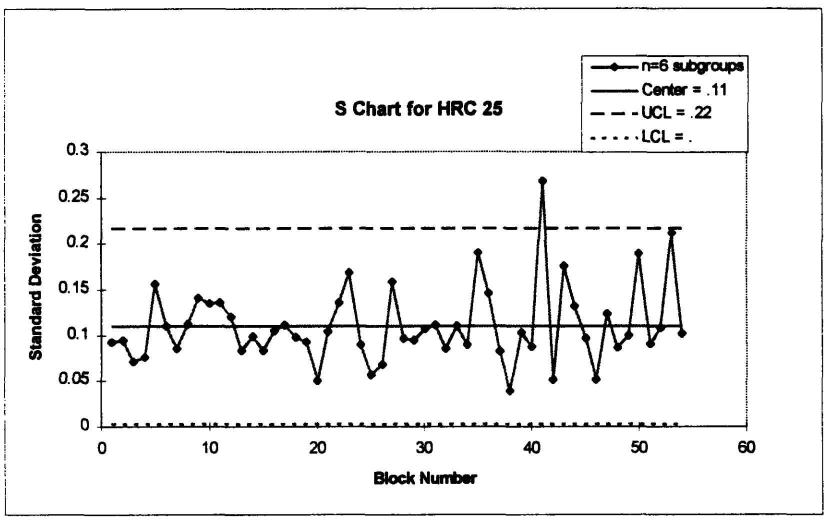
However, Instron currently offers 120 different major part numbers, 60 for the regular scales ('A', 'B', 'C') and 60 for the superficial scales ('T','N'). In addition, the firm offers minor hardness scales and levels with extremely low usage (e.g. 3 per year) and Large grade blocks of 3 hardness levels HRC 25, 45 and 63.

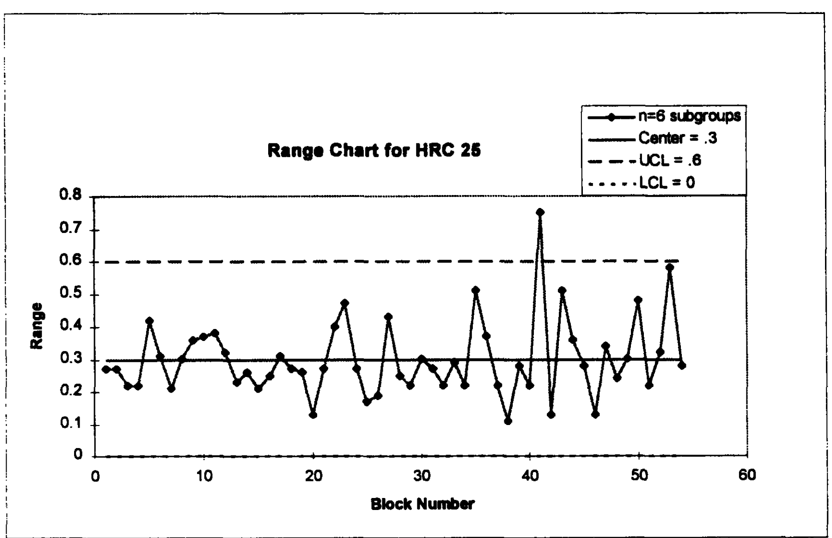
For three charts ($\overline{X}$, s and R) per part number, up to 360 individual charts would be required. Someone would also have to look at them all for diagnosis purposes.

The author maintains that statistical control for the general process can be established on the basis of fewer, higher usage part numbers.

Figure 8-2: Examples of $\overline{X}$, s and R control charts







8.6.3 Chart set-up for small lot sizes/short runs

The average manufacturing lot size for block production ranges from 50 to 200 blocks, with a lot size of 100 for a particular hardness level being the most common. Shewhart traditionally advocated setting up charts and calculating control limits with 100 to 125 data points [31]. Devor suggests: "As a good rule of thumb, one should select 25 to 50 samples to provide a solid basis for the initiation of control charts"[14].

Typically, control charts are set-up in order to achieve an initial basis of statistical control from which to diagnose out-of-control conditions [14]. This set-up provides initial estimation of the process mean and variability for the determination of control limits. "Once control is established the charts may be employed to monitor the future behavior of the process" [14].

8.6.4 Inventory control and traceability

The five steps of SPC of Table 8.1 require that the transient flow of blocks through the process and into calibration measurement maintains its time-based integrity e.g. blocks are processed in *sequential order*. Refer the process control viewpoint, Fig. 8-1. Otherwise, it becomes more difficult to go back into the process and to recreate the history of a particular production lot for the purpose of investigating and removing common systemic causes. As a result, it is imperative that work-in-process is controlled with disciplined conformance to clear inventory control policies. For example, uncalibrated blocks are separated by production lot and sequenced on a first-in-first-out (FIFO) basis.

To allow diagnosis and corrective action, information regarding the parameters of manufacture must also be compiled and maintained with traceability to the production lot (or ideally to an individual block). Examples of such information include:

- Material test certificates on composition, purity, grain size
- Heat treating process records and furnace temperature traces
- In-process hardness measurements
- Machine parameters in finishing, for example grinding speed, feed and depth of grind

Such traceability requires process documents to be related to process lot numbers and these in turn to be linked to serial numbers of individual blocks contained in them. The complexity and quantity of information must be efficiently managed.

8.6.5 Feedback Delay and Information Quality for Corrective Action

Fig. 8-1 depicts the feedback perspective of statistical process control of the block manufacturing process. It is assumed that traceability to the systemic cause becomes more difficult as time passes. It is more difficult recreate history as the interval between present and the distant past at which the systemic event took place becomes longer.

There are several delays that will influence the feedback response:

- The cycle time of the process itself, Tp
- The time in which the blocks remain in inventory prior to measurement, T_i
- The delay associated with differentiating tester variability from block/process variability, as well as completing the observation-evaluation-diagnosis-decision-implementation feedback loop, T_f.

All of these delays undermine the ability and effectiveness of corrective action and hence decrease the utility of SPC application as a whole. These delays are enhanced by the number of operations, locations, and parties along process chain.

8.6.6 Tester variation masks block variation

The tester is part of the process and has its own sources of systemic and random variation. The goal is to measure and control the variation in the block process. With the large portion of measurement error attributed to the calibration tester, it becomes difficult to discern if an out-of-control condition is to be attributed to block or tester variation.

As long as the standard deviations attributed to the tester is a significant fraction (e.g. greater than 25%) of the standard deviations attributed to the block, we cannot distinguish one source of an out-of-control condition from the other. A special process control strategy must be thus devised to leverage the enablers and to deal with the challenges.

Chapter 9 A Comprehensive SPC Strategy

A comprehensive approach for achieving statistical process control is proposed in light of the stated challenges. This strategy is treated independent of a formal cost/benefit analysis in the context of the competitive business strategy.

9.1 Process optimization using designed experiments (DOE) and deadweight measurement data.

This approach to process improvement for the manufacture of blocks is performed on the outset as a focused project of limited duration. As such, it allows for a larger number of indentations to be performed on the blocks. Thus, sampling error in characterizing block and process variation is reduced.

The experimentation for process optimization would have three primary goals:
- Minimize the block variation (non-uniformity) at each hardness
- Define a set of process parameters and procedures for robustness[12] to hardness variation
- Improve controllability of the nominal block hardness through tempering or otherwise

A fourth goal to relate the behavior of block variation to hardness level in the form of a verified model can be considered. This would allow the scope of future process control efforts to be reduced.

Because of the focused nature of this process optimization activities, it is possible to parse and characterize block variation (s_{block}) from tester variation (s_{tester}) by leveraging improved tester technologies, such as the deadweight tester from NIST. Refer to the parsing procedure using the components of variances models of Chapter 3. The deadweight tester would need to provide a statistically significant quantity of block measurements in order to sufficiently reduce sampling error. As a minimum n=15 measurements per block are required, whereas n=30 measurements is ideal.

As an output of the designed experiments, a set of parameters and procedures for machine, material, methodology would be defined with a identification of which parameters carry more weight than others. For instance, parameters in material may have more influence on block variation than changes in heat treating parameters. As many of the known process parameters and methods must be documented as a foundation for maintaining statistical process control through constancy.

It should be noted that Design for Robustness methods, e.g. Taguchi experiments, require purposeful parameter changes (high and low) from the current operating points in order to detect how the block variation is influenced [21]. For example, an alternate material supplier with measured changes in composition may be tested. In some cases, this will result in blocks of

---

[12] Robustness is defined as the process' resistance to disturbances in material, machine, methodology, environment e.g. small fluctuations in process parameters [21]

increased variability that do not meet customer specifications and hence cannot be sold. The cost or risk of non-revenue yielding experiments would need to be considered.

This study was in part the result of such process optimization experimentation. Chapter 3 describes how the feedback of NIST deadweight measurement data on Large blocks of higher quality was used to isolate block variation from tester variation in order to study the causes of block uniformity.

The parameter results of this DOE process should be verified with confirmation experiments.

9.2 Focus on a select few, high-usage part numbers and *Copy Exactly*

Several types of scales and nominal hardness levels within them are subject to significantly higher customer demand than other part numbers of the 120 offered. Refer to Section 8.6.2. The author proposes that a select few Rockwell hardnesses/part numbers are chosen for achieving control of their batch processes. The part numbers are to be chosen based on the following characteristics:

- high-usage
- distributed over the different tester scales
- distributed over the equivalent hardness levels
- representative of the different material and process alternatives.

For example, 3 superficials, 3 regulars, of low to high hardness, brass and steel would result in 24 part numbers. This would require approximately 24 x 3 = 72 control charts or an 80% reduction in chart quantity (See Secion 8.6.2).

The parameters and methods for the in-between part numbers and hardnesses not directly measured and for which control is not explicitly determined must be continuously adjusted and indirectly monitored to reflect those of the focus part number that are subject to SPC. The author uses a term *Copy Exactly* developed by the Intel Corporation [41] for starting up new production lines as mirror copies of the pilot processes. The discipline of the *Copy Exactly* method implies that every minor detail of the source process is duplicated, even aspects that are deemed as inferior or are held to be systemic causes of variation. Only in this manner, can a substantial probability exist for extending process control from representative focus blocks to those not directly monitored.

9.3 Automate chart-making and chart pattern diagnosis

Even though the number of charts has been reduced, manual creation and interpretation of 72 charts represents a considerable workload for skilled production associates. Hence, it is imperative that the data collection, data management, control chart creation and, in part, diagnosis of control charts be computerized.

Commercially available database and SPC software packages readily accept the ASCII data that can be downloaded from the respective standards testers. [26]. Diagnosis aids, though not a substitute for SPC training, can be helpful in detecting patterns and out-of-control conditions in the control charts that are evidence of systemic causes variation. The equipment feasibility for automation of data collection and control charting was confirmed by the author in a pilot tester set-up at Instron.

9.4 Gradual chart set-up with reference to specification limits and chart continuation between process batches

The control chart set-up with sufficient number of initial data points in a low volume environment is in part aided by focusing on the higher usage part numbers [See Section 9.2]. However, we are still dealing with maximum *calibration lots* 40 blocks for *process lots of 100 to 200 blocks*.[13] Every block measurement subgroup of n = 6 measurements is to be charted.

The limited batch sizes (short runs) require that *charts are transferred from lot-to-lot* within the same hardness level and scale. In doing so, it is important to designate the calibration and process lots in order to support investigation for special causes. Much can be learned by recognizing shifts in variability (s and R) that occur between either calibration or process batches. For example, mean or variation shifts between calibration lots for blocks stemming from the same process lots would indicate that there may be special causes associated with measurement noise in the tester system. This chart transfer between lots and designation of block serial numbers to process lot number is easily accomplished within a computerized database system [See Section 9.3].

Hence, a gradual chart set-up procedure is required over the first two or three lots in order to ascertain reasonable estimates of the grand means, average range, average standard deviation and the respective control limits on the $\overline{X}$, R and s charts. This challenge also exists for EWMA and CUSUM charts for low sample quantities [14]. The chart is started with trial limits based on 'typical' values from individual samples reflecting the amount of variation that could be expected; for instance using the values found in the process optimization studies [18].

In contrast to most applications, it is recommended to *plot the specification limits* on the $\overline{X}$ and R control charts, particularly in the early phases of chart set-up. As was cited in Section 8.5.3 the summary statistics of the n=6 subgroups charted can be directly compared to customer specifications e.g allowable range or average nominal hardness for n=6 measurements. This will help determine if isolated extreme points are cause for immediate process investigation action. For instance, the $\overline{X}$-chart depicted in Figure 8-2 shows two immediate extreme points beyond the 3-sigma upper control limit, although these do not exceed the allowable tolerance range on nominal block hardness specified in Table 1.1 for HRC 25.

The benefit of showing specification limits is that the $\overline{X}$ and R control charts also serve as immediate reports on product performance and quality levels.

---

[13] A process batch may incorporate approximately 100-200 blocks that are inventoried after final polishing. A calibration lot may on the other hand only range from 5 to 40 blocks.

Thus, the chart *control limits* (UCL and LCL) are not regarded as fully functional until at least two process lots have been calibrated and charted. The data is compared with the trial control limits to see whether the variation is stable and appears to come only from common causes. Investigation of special causes may begin right in the outset. In the early phases, it is more effective to look for patterns or trends in the data points as signals of systemic variation as opposed determining violations of control limits. This is the procedure employed for EWMA charts where 'the detection of shifts is based primarily on trends in the data' [14].

9.5 Separate control charts by tester and assign testers to hardness scales

Statistical control and special cause investigation can be achieved only if each of the 24 focus Rockwell scales and hardnesses (of Section 9.2) to be charted is assigned to a specific tester system in the Standard laboratory. This constraint is required since the tester variation is part of the total measurement variation seen on the control charts. In order to develop a functional control chart history between process lots, the measurement variation must be characterized with a tester variation that is held as constant as possible over time.

Another reason for matching the hardness scales and corresponding charts to a specific tester, is that the commercial tester variation is dependent on hardness level [Refer to Section 2.3.4]. It is currently not known if the correlation behavior of tester variation to hardness level is the same *between* testers. Therefore, mixing of Standards testers for a given hardness chart over lots may make the chart control limits meaningless for continuously changing tester variation.

There are currently six available Standards testers. Due to the even partition of part numbers between superficial and regular scales, three testers can be assigned for Regular scales and three for the Superficial scales with balanced distribution of representative hardnesses, materials and processes. This would results in approximately four major part numbers per tester of three charts ($\bar{X}$, R and s).

Assigning the block part numbers to individual testers in this manner carries one fundamental requirement: All six testers must be maintained in proper operating condition. Also immediate response by a qualified technician if a tester failure or a tester-attributed systemic cause has been detected is required.

9.6 Parse-out tester variation using deadweight tester measurement feedback

If an out-of-control condition is detected, the first step is to determine if the tester system used to measure the blocks in question is in a state of statistical control before engaging on the more arduous task of investigating the block process.

In order to determine the control/variation condition of the tester when an out-of control condition during block calibration is detected, reference blocks and a Rockwell tester of considerably lower total variation (Large grade blocks and a deadweight tester) can be used as an effective leverage tool to parse tester variation from block variation. The following scenario determines if the tester variation has gone out-of-control and become excessive due to special causes:

Process for Detection of Special Cause Sources:

1. Instron supplies 'best' blocks of lowest non-uniformity available (e.g. Large grade) to NIST; at least one block is supplied for each of the part number/hardness types (qty. 24) under SPC focus along with a *matched indenter* for each.
2. NIST measures the blocks using the low variation deadweight tester with the paired indenter provided. A statistically significant number of measurements of n=15 to 30 indentations is required to attain good estimates of block variation.
3. NIST returns the measured block, the matched indenter and corresponding measurement data to Instron.
4. The Instron Standards tester that detected the apparent out-of-control condition is used to re-measure the same best block using the same matched indenter. Again, a statistically significant number of measurements (n = 15 to 30) at the focus hardness level serves to provide a sound variation estimate.
5. The components of variance model (See Fig. 3-2 and Equation 8) is applied in order to parse the Standards tester variation (s_{tester}) from block variation. Refer to Figure 9-3.
6. The s_{tester} is compared to previous levels determined by this method (initially during process optimization 9.1) to evaluate if s_{tester} is excessive. A separate s-chart for tester variation must thus be maintained. Each s_{tester}-chart is linked to each specific Standards tester (See 9.5)
7. If the tester variation s_{tester} has not significantly changed (e.g. using a chi-squared test), the out-of-control condition may be attributed to the test block. If the special cause variation has been attributed to the test blocks, investigation into the block process for special causes by the feedback loop of Figure 8.1 is warranted.
8. Otherwise, the special cause can be attributed to the Standards tester. The special cause investigation can begin with indenter used to calibrate the production blocks.

This special cause detection process for the tester systems is depicted in Figure 9-3.

Since a test block can only fit up to 120 to 150 indentations on average, approximately 5 sets of internal tester verifications of n = 25 measurements are possible per special deadweight calibrated block. The larger number of measurements are crucial to minimize sampling error and give better results of the chi-squared test for equal variances [15].

The tester diagnostic process of Figure 9-3 is performed *reactively* to an out-of-control condition on the control charts. *Proactive tester monitoring* entails performing the tester check at frequent, regular intervals, for instance daily or weekly to establish the tester control condition in advance. Given the substantial cost of the test blocks that are calibrated by a (NIST) deadweight tester, the proactive approach does not seem feasible. The reactive approach requires a higher level of operator responsiveness during the regular production routine in order to immediately evaluate the tester control condition when a control chart determines an out-of-control condition.

The feedback process defined above with deadweight calibration by NIST provides an added benefit: It serves as a vehicle for ensuring continually updated cross-reference of commercial standards testers to the national standard, developed and maintained by NIST.

Production calibration using a deadweight tester is assumed not feasible

The reader may ask: Why not bypass the complex strategy of variation parsing by using a deadweight tester to calibrate test blocks in production ? Wilson Instruments formerly operated a deadweight tester in order to maintain their commercial hardness standard at a time when the commercial Rockwell tester technologies were not as precise. Wilson Instruments previously determined that direct production calibration with a deadweight tester is not to feasible due to the additional processing time and cost required for deadweight tester operation. [13].

Consideration of an alternative strategy for isolating special causes

An alternative strategy for determining if the Standards tester is the main source of the measured out-of-control condition was considered: Suppose that three different Standards testers were used to calibrate a lot of production test blocks of the same part number/hardness level *at the same time*. It is intuitive that it is less probable that two or three testers would go out-of-control (develop excess tester variation) at the same instant in time. Hence, if only one out of three testers signals an out-of-control condition, probability strongly suggests that the special cause is attributable to the block process.

The author abandoned this alternative strategy for the following reasons:

- In order to provide sufficient sampling quantity to detect a condition of out-of-control, each Standards tester would have to calibrate a significant number of blocks, e.g. a minimum of 25 to 30, each time the particular hardness level/part number required calibration due to demand. Calibration quantities of 50 to 90 blocks at a single calibration interval are prohibit maintaining a flexible and cost-effective policy for finished, calibrated block inventory.

- It is necessary to match each control chart to a specific tester in order to maintain constancy in tester contribution to charted process variation, as discussed in 9.5. The minimum quantity of required control charts is thus magnified to 216 charts (3 charts x 3 testers x 24 part numbers).

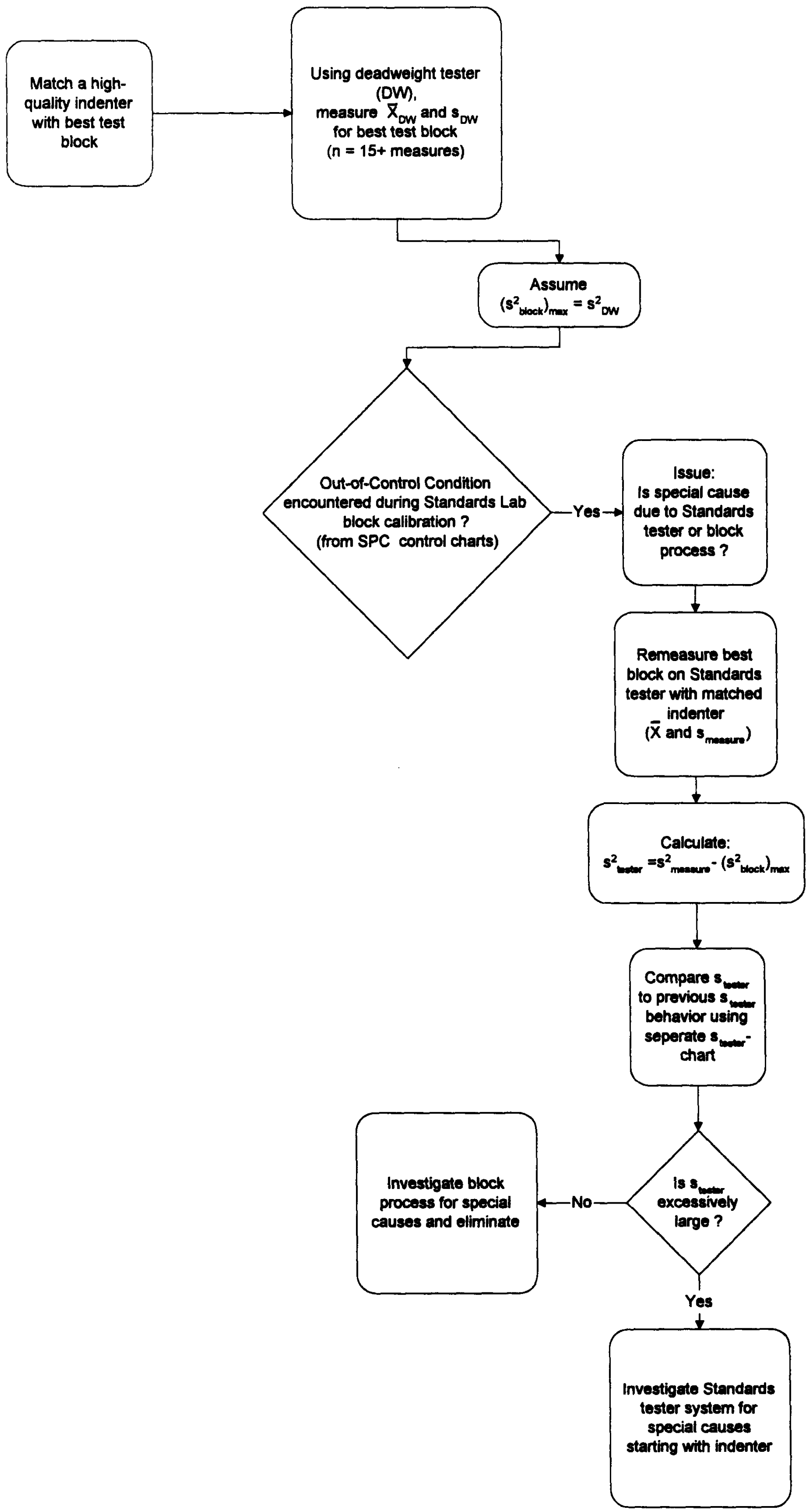
- Every time a part number requires calibration for production demand, all three Standards testers must be in full operational condition. This does not allow for convenient rotational scheduling of special cause tester investigations and regular tester maintenance.

9.7 Use tighter internal specification limits or determine C_R and Cpk $\geq$ 1.5

The tester technology of the customer base may be significantly inferior to the Instron standards testers from a tester variation standpoint due to lower state of technology or maintenance and control conditions.

Let's take the case, in which the Instron tester demonstrates a lower measurement variation s_x than the total measurement variation s_y measured by the customer. Refer to process diagram of Figure 8.1. If Instron references the customers' specification limits on range (ASTM E-18) and ships a block that measured just within those range spec limits, it is probable that the customer will measure a range that exceeds their specification range limit due the higher tester variation of

Figure 9-3: Process Diagram for SPC Detection of Special Causes in Test Block Production



the customer's tester. They will more than likely perceive that the block has exceeded the allowable range limits established by the standards institutions [See Table 1.1] Thus, the customer will deem the test block to be unacceptable.

This highlights the notion that both specification limits and control limits can and probably should occur on the same chart for hardness test blocks [See Sections 8.5.3 & 9.4]. One approach is to place *tighter internal specification limits* on the control charts.

Another approach to the problem is to apply the capability metrics Cpk and C_R for nominal hardness and range, respectively. These are useful metrics to gage the likelihood of exceeding the customer's requirements/specification limits.

The example of Section 7.2 showed that a C_R value of 1 exists when the maximum range just equals the specification *using the Standards tester*. To compensate for the variation of the customer's testers the critical value of C_R (and similarly Cpk) can be increased in order to determine if the process is currently meeting customer requirements. From the capability metric equations (33) and (34), it can be seen that the critical values can be scaled in direct proportion to the ratio of customer (Y) to standards measurement (X) variation. For example, for a s_Y/s_X of 150% the critical capabilities are 1.5. Hence, as long as the process exhibits C_R and Cpk grater than 1.5, Instron can be assured that the customer requirements are satisfied.

It is to be noted however, that the measurement variation must still be random in nature. Signs of systemic causes (e.g. patterns) may be early signals of a deteriorating control; the out-of-control progression of blocks can continue right beyond the specification limits.

9.8 Computerized Process Traceability and Inventory Management

In order to minimize the delay associated with the root-cause investigation when/if an out-of-control condition is encountered, the author advocates using *bar-codes* to link the block serial number to the measurement data and its process history.

Bar-code stickers can be placed on the bag container of a test block after its has been polished and is placed into inventory waiting to be calibrated to a particular scale and hardness. Note that simple bar code stickers cannot be placed on the block itself without interfering with the hardness measurement. The bar code would contain the serial number and a letter digit representing one of the possible 'blank' part numbers representative of the particular heat treating process.

Each tester can be fitted with a hand operated bar-code reader or wand, wired in parallel to the Rockwell tester and the data acquisition system. As such, each hardness measurement data stream can be linked with a part number from the calibration technicians bar-code scan.

This bar-code set-up was successfully piloted by the author for the data taken in this study. In addition, Instron has in the past engineered and supplied automated tensile testers that operate under this same model.

The part number identified as representative of an out-of-control condition could thus be identified by the SPC control charting program [27]. A simple database search would identify the process lot number of which the part number belonged. The process lot number can quickly link to the documentation of process history described in Section 8.6.4.

All blocks should be processed and inventoried according to process lot grouping and time sequence to facilitate transient trend diagnosis. [See Section 8.6.4]. A simple first-in-first-out (FIFO) policy goes a long way to maintaining processing sequence.

The author also points out that the information systems defined for the purpose of statistical process control also provide complementary benefits with little extra work. These complements include:

- database access of block calibration measurements to support customer service in helping customers with tester validation issues.
- computerized calibration certificates to eliminate transcription errors by technicians and additional inspection of summary statistics and individual measurements [27].
- customized calibration certificates to customer name. A whole market potential exists for specialized and replacement certificates of reference standards.
- rapid and reliable inventory auditing and usage tracking by part number. Outgoing quantities can easily be tracked with a bar-code scan of the block container. End of term inventory counts by part number can be accomplished by bar-code scan.

Part IV Organizational and Management Issues

Chapter 10 Organizational and Management Issues to Introducing Statistical Methods

If what we know about our processes can't be expressed in numbers, we don't know much about them.
If we don't know much about them, we can't control them.
If we can't control them, we can't compete.[14]

10.1 Statistical Methods for Product and Process Improvement

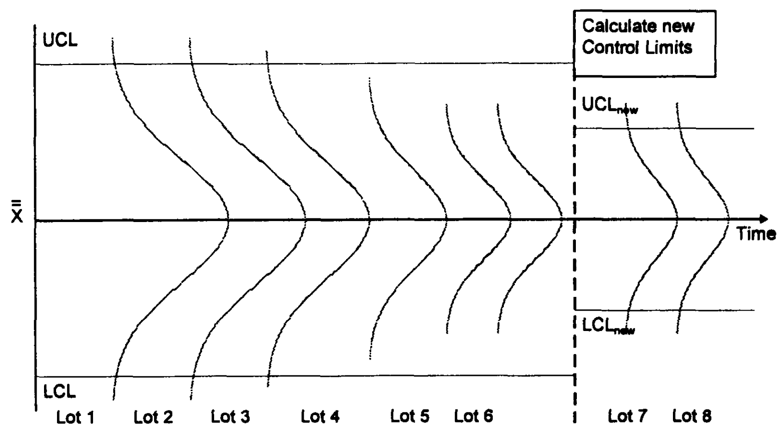
The statistical methods described in this study are tools that aid in the search for product and process improvements. Improvement entails continuously addressing the sources of variation as they are detected. Control charting serves to detect potential problems so that action can be taken BEFORE the *quality of the product* is compromised.

Due to the close relationships between the components that make up a Rockwell tester system [Refer to Figure 2-1] and the variation contributors to the Reference Standard [Figure 3-1], the *product* subject to control and improvement extends beyond the test blocks. The control charting also monitors the state of the standards testers; these in turn influence the quality of indenters, production testers and calibration information provided to the customer. It was also demonstrated that achieving statistical control of the block manufacturing process requires attaining control of the measurement standards testers, as part of it. The Calibration Capability links the improvements in the state of block technology and the parallel improvements in the tester system and its subcomponents in a metric of combined result.

The process for diagnosing, investigating and eliminating sources of variation was introduced in Chapter 8 as a feedback loop of SPC. This process is often termed independently as SPS, *statistical problem solving.* Over time, the iteration of this root-cause elimination, in conjunction with off-line process optimization studies, should result in a decrease in overall variation. As a result, the control limits of the $\overline{X}$, R and s control charts require adjustment. Refer to Figure 10-1 below. Note that the appropriate time for adjustment is signaled by the control chart diagnosis rule number 8: When 15 successive points fall within +/- one standard deviation of the centerline [14]. The decision to decrease the specification limits with regard to customer requirements and expectations should be treated as an independent one.

[14] Dr. Mikel Harry, Research Director at Motorola University [18]

Figure 10-1: The effect of process improvement activities on SPC control charts



10.2 SPC as an integral part of the business strategy

SPC clearly requires investment in new skills and in the early phases will require much resource commitment. In order to maintain constancy of purpose and justify the up-front investment, it is imperative that management can clearly relate the SPC activities to the business strategy. 'The purpose of justification is to get commitment from top management' [29].

The author imparts several viewpoints on how SPC can be used as a competitive weapon:

- *The competition is using SPC.* Yamamoto, a leading Japanese competitor, uses SPC for the production of their hardness testing systems [3].

- *Customers are using SPC.* Instron routinely supplies testers with special SPC features that are supported by computerized information systems. In addition, the firms of Instron's customer list include world-class manufacturers that are known for their application of SPC. Developing the in-house capability for SPC can serve to better understand customer's needs for the measurement systems in the future. This learning can be translated into improved product offerings.

- *SPC improves process and product knowledge in order to stay ahead of the competition.* The learning that is gained through the investigation and removal of causes of variation can result in improved quality products. The learning will be not only be with regard to the test block processes, but the sources of variation in the tester system as well.

- *Improved product quality results in higher market share.* This assumption holds true if:
  (a) customers can adequately perceive the product improvement relative to the competition
  (b) improvements do not sacrifice the equilibrium price/performance relationship
  The product performance differentiation of (a) can be accomplished by empowering and educating the customer base with the Calibration Capability metric and the statistical underpinnings.

The costs associated with achieving the appropriate price/performance relationship can be *reduced* if adequate regard is given to overall operations costs.

- *Cost reduction.* The learning about the internal production processes can be initially translated into cost reductions by eliminating redundancies and improving efficiency/throughput. This was the result of the process optimization work leading up to this study. The view of cost reduction opportunities must include those costs that aren't readily apparent and that don't appear on internal accounting records. Refer to Table 10.1 that summarizes the *costs of quality* from Devor [14] and Shecter [25], respectively, that need to be assessed in order to confirm positive SPC benefit. It must be evaluated if the prevention and appraisal costs can offset the costs of internal and external failures.

- *Barriers to entry.* The strategy for statistical control encompasses the global measurement system of blocks, tester, indenters and operator influence. A potential market entrant cannot viably supply only one element of the system, as they do not control the necessary elements to achieve process control. It follows that without control, capability is vague and quality is happenstance.

Table 10.1: Costs of Quality [14, 25]

Cost Of Quality	Cost Items	General Definition
Prevention	Quality planning	Maintaining a quality control system
	Training and motivation	
	Process planning	
	Vendor selection	
	Design review	
	Parts selection	
	Qualification	
	Reliability analysis	
	Control Charting[15]	
	Root-cause investigation and elimination[16]	
Appraisal	All inspection: Incoming, in-process, and final	Maintaining a quality assurance system
	All tests: Incoming, in-process, and final	
	Quality audit	
	Calibration	
Internal Failure	Rework	Manufacturing losses , scrap and rework
	Retest	
	Scrap	
	Troubleshooting	
	Failure analysis	
Extra Operations	Corrective action costs	
	Excess inventory costs	
External Failure	Warranty	Warranty, repair, customer, and product service
	Customer complaints	
	Customer returns	
	Added services for field corrections	
	Lost customers	

[15] Cost of setting up, maintaining and diagnosing control charts was added by author as an individual item of quality control costs.

[16] Cost of root-cause elimination was added by the author as a prevention item from a proactive perspective, as opposed to internal failure, corrective action, which is reactive.

10.3 Data-driven Decision Making

The statistical methods for capability and control require collection and analysis of process data. The return of these efforts is the ability for production associates and management to make decisions that are based on more factual data than judgment and inference.

Because managers cannot be immersed in the routine activities of block production or calibration, there exists a natural tendency to focus on that data that is readily *available* (e.g. occasional probing into the work of subordinates) or to *anchor* on sample data from an unusual event or crisis. The concepts of *availability* and *anchoring* are familiar hurdles to objective risk assessment [36].

The proposed control charts plot every block that is calibrated and they allow for specification limits to be placed on them directly [See Section 9.4]. In conjunction with the added calculations of the capability indices, the control charts provide a concise status report on the product and process performance that also show the transient changes over time. As a result, the manager gains a more complete and representative view into the state of the production system. This improved view features a sense of average performance, as well as the bounds of variation.

10.4 Other External Benefits to SPC

The introduction of methods and systems in order to conduct statistical process control for Instron's test block production carry external benefits of considerable merit.

The computerized data collection and management systems for control charting purposes also allow for:

- customer calibration certificates to be automatically generated without transcription or calculation errors [27]. Secondary inspection of certificates can thus be avoided. Customized certificates by customer name would also provide added value to the reference standard.
- a database that can be accessed by customer support personnel to better service customers with tester validation issues.
- rapid inventory and product usage tracking for production scheduling and capacity planning.

These benefits map into the costs of quality of Table 1.1 as cost reducers.

10.5 Probabilistic Significance Levels relate to Acceptable Customer Quality

Calibration capability is assessed based on a level of significance greater than 99% for a test statistic $z = 3$ [Equation (31)]. This means that less than 1% of the time the assessment that the parent average and dependent average differ by less than block half-tolerance may be wrong (Type I error). Control chart limits are similarily based on a +/- 3 sigma level, representing that 99.7% of the block variation is expected to lie within the control limit value.

It is important that management and production associates understand the meaning of these inferential probability assessments. The significance levels and choice of the critical test statistic, implicitly define what is the *acceptable quality* (number of defects) that may pass to the customer. The author thus has taken some liberty in choosing significance levels deemed to be adequately conservative with regard to industry benchmarks and that are commensurate of other known statistical metrics e.g. Cpk.

In addition, these inferential assessments are made on the basis of small samples (e.g. n=6) whose estimates will be subject to sampling error. The role of sampling error and the assumptions made in the representative metric must thus also be considered in the 'risk' assessment. In order to compensate for sampling error the capability metrics can be used in there existing form with the modified requirement that the minimum allowable capability is greater than 1 e.g. 1.5.

10.6 Customer Focus of the Calibration Capability

The systems perspective embodied in the Calibration Capability (Cc) index provides a perspective with particular attention to the customer's environment and needs. [See global system model, Figure 3-1] For the block to maximize its customer utility as a reference standard, the degree of variation not only in the calibrating tester, but also the customer tester must be sufficiently controlled. To make the Cc index work, sufficient sample data from the customer measurement system(s) must be gathered.

It follows that through close customer interaction over time customers will become informed about the Calibration Capability index and what it represents. The customer may be provided with sufficient data on the calibration certificate to perform the Cc evaluation themselves. The customer's quality perception not only of the blocks, but also of their tester systems and that of the standardizing laboratory could thus be formed by the Cc index.

The Cc index may become the yardstick by which the customer is empowered to evaluate suppliers of the hardness reference standard. Those competitors who in the past have been able to mask their true block and tester performance, will be forced to uncover the true nature of product performance. As a leader in Rockwell testing equipment, Instron can better present the value of their measurement systems and reference standard products in relation to the competition. The Cc index may thus serve as a tool for gaining competitive advantage.

For Instron the data collection process associated with the customer's tester performance provides an opportunity to learn about the customers environment and needs. The 'Customer In' transfer of information from the customer's environment is strongly advocated by Total Quality Management principles for product and business improvement [24].

10.7 SPC requires processes for corrective action

The feedback perspective for SPC presented in 8.3 highlighted the importance of diagnosing, investigating and correcting special causes of variation. This feedback loop in itself is a process that must be defined for the Instron production chain.

The field of TQM provides useful analytic and communication tools within a process framework, called the *7 Steps process*, to aid in the team-based root-cause investigation [24]. Refer to the other tools presented earlier in Table 8.1. An example of such a tool, the Ishikawa diagram, applied to the block manufacturing process (surface grinding defects) is depicted in Appendix I.

It is important not to lose regard for applying engineering knowledge and basic science in the investigation of special causes. Data-driven decision making must be balanced with established scientific models and experience that can be readily applied. "There is no substitute, statistical or otherwise, for firsthand knowledge of the process that needs to be controlled" [34].

Communication and cooperation between the Standards laboratory and the production personnel is fundamental to this corrective feedback. The communication must be efficient yet effective. Instron is faced with the added challenge in that block machining, heat treating and measurement/calibration are performed in three distinct locations. Many case examples in industry show that co-location of the process activities is a key communication enabler, particularly as the removal of organizational boundaries help remove subgroup loyalties and affinities. In addition, a closer business partnerships that extends beyond an arms-length contract may be required with external vendors.

Standardization of process operations (See following Section 10.8) tasks demands communication between production shifts. At NUMMI[17] one hour is scheduled between shifts to accommodate meetings for the two shift teams to evaluate tasks for common standardization [40].

The role of management involvement in the improvement activities is underscored. "Only a relatively small number of process troubles - industrial experience suggests about 15% - are correctable locally by people directly connected with the operation; the majority are correctable only by management action on the system" [18]. The real leverage thus lies in arming management with the skills in statistical problem solving [32].

10.8 The Discipline of Statistical Control

The SPC strategy of Chapter 9 presents the importance of establishing standardized process parameters and procedures in sufficient detail to ensure that they can be consistently repeated. This entails documentation of how work is performed in the form of process and operation procedures. These procedures must also allow for continual updates to reflect the removal of special cause sources. "Once tasks have been standardized they can be improved" [40]; root-causes can be more easily investigated and solved since they are subject to less noise.
These standardized procedures also serve as a communication tool to pass information up and down the process chain, as well as up and down the structures for managerial decision-making. Finally standardized procedures serve as a training tool for new production associates and engineers. Such standardization of work, information flows, and business processes is the cornerstone of ISO 9000 [37].

[17] The New United Motor Manufacturing auto assembly plant in Fremont, CA is a GM-Toyota joint venture with a work system based on the Toyota Production System.

Establishing standardized procedures in itself is not sufficient, although a first step. Achieving and maintaining statistical control demands *strict conformance* to operational procedures and conformance requires the *discipline* of individuals. Management must therefore support that discipline by stressing the importance of consistency in work even when at times that it may conflict with production scheduling pressures.

The strategy for achieving statistical control of the test block production process [See Section 9.2] is based on the ability to 'copy exactly' the process methods and parameters used for the selected focus part numbers to the other part numbers. In order to 'copy exactly' the process procedures for the focus blocks must be well documented. Discipline is required in carrying over practices in explicit detail; even those that may have apparent deficiencies or bear apparent special causes of variation. The systemic interaction of these shortcomings with other parameters are unknown or cannot be sufficiently modeled and tested. Management must therefore be rigid in emphasizing this discipline in order to make the 'copy exactly' strategy work.

In addition, management must motivate against haphazard adjustment of the process. Devor et al use the term "over-control" for the case when operators prematurely adjust the process while it may still be in control [14]. Over-control can occur when associates falsely interpret a natural fluctuation in a machine or process as the effect of a special cause of systemic variation.

10.9 Standardization of tasks requires adjustment of production roles

Klein points out that the standardization of tasks that is so beneficial in controlling variation in the (block) process also eliminates variability in the way production associates may approach and think about their daily tasks [40]. Thus, standardization removes a certain degree of *autonomy* and *variety* in the job functions of associates; key aspects to individual job satisfaction are thereby eliminated.

One solution to the inherent conflict between process consistency and the individual's job role needs is employed by Toyota [40]. Operators are involved in controlling the *design of tasks* together with other operators as a team. As such, a collective autonomy is generated in determining work methods, although flexibility in individual task execution is strictly limited. Variety on the job is provided by the activities associated with improvement problem-solving. However, improvements are collectively made by teams or Kaizens in a selective and structured manner of decision-making that involves management.

10.10 Using the capability metrics to drive improvement

When applied correctly, the three capability metrics (Cc, Cpk and C_R) are both encompassing and founded in sufficient detail to serve as meaningful and brutally objective assessments of overall product and process performance. They are thus suited for management to track improvement and to use these metrics as goals for the future desired state of the process and organization.

The process capability index, Cpk, is being successfully used by world-class manufacturers to continually track and motivate process improvement, including Ford, Kodak and Rockwell International.

The capability metrics are also well suited for use with a TQM management approach, known as *hoshin management* or *policy deployment*. Hoshin management uses explicitly defined goals which are used to quickly align all people, tasks and functions within the organization [24]. Hoshins are statements of the desired outcome for the year combined with the definition of means of accomplishing the desired outcome and the metrics for accomplishment.

Hoshin = statement of the desired outcome for the year
+ focused means
+ metrics to measure progress
+ target value of the metric
+ deadline date

The hoshin is furthermore translated into the roles of the respective functions. For instance, product development for tester improvement, purchasing for vendor improvements etc. Hoshin management has another purpose: "it forces managers to run the Plan-Do-Check-Act cycle themselves as a part of their daily job" [24].

The capability metrics Cc, Cpk and C_R are well suited for use as hoshin metrics if their founding data is objectively gathered. Even the simple hoshin goals are subject to the influences of bad or tainted data. It must be reemphasized that management cannot gage capability without having confirmed the condition of statistical control.

10.11 Planning for Changes in Introducing SPC

Before answering the question of whether or not to commit to the pursuit of SPC, Instron management needs to define a clear perspective of 'what we are getting into'. The purpose of such a planning activity is to ensure that resources have been adequately accommodated, challenges are anticipated and change processes are initiated. For instance before collecting any SPC data, it is important to ensure satisfactory training of statistical methods for the right individuals has been accommodated. Kiemele et al emphasize the importance of ensuring that 'an environment for improvement' has been created' [18]. While there exists no single magical recipe for the adoption of effective SPC and continuous improvement, there is sufficient literature on case studies and lessons learned to be gained. The necessary changes will also involve the learning by management itself, as it will need to converse and decide on issues of a statistical nature.

New levels of overall performance require new methods to approach the overall problem. The author maintains that the statistical methods proposed herein provide the basis for dealing with the *technical* aspects for solving the performance problem. However, these statistical methods demand new skills and new ways of conducting work that lie at the heart of the Instron's business processes. For these methods to be effective management must also provide the means and leadership for *learning* and *adaptive change* required of the organization [35].

The scope of the many of the technical capabilities required for implementation of the SPC currently exist within Instron's organization; many of the individuals that possess the particular expertise are however committed to other departments and products. For instance, trouble-shooting for systemic variation in a standards tester will require the expertise of an experienced tester technician from the production line. The computer system for SPC data management requires know-how harbored by the IT department. Instron electrical engineers have also previously set-up automated SPC systems for customers.

Due to the interrelated nature of the tester and block sources of variation, the continuous improvement activities for producing blocks cannot be an isolated responsibility of those parties associated with block production and calibration. The entire Instron Rockwell hardness organization must therefore be aligned in these efforts.

An effective tool for planning for organizational change was developed through the LFM program at MIT, called the 'Matrix of Change' [38]. This tool engages management to identify its current state and capabilities and to identify the new state and capabilities required or desired. Relationships between existing practices can be made to the target practices of the future in order to predict the difficulty and challenges and to anticipate unsuspecting interactions.

10.12 Relationship with NIST and the national standard

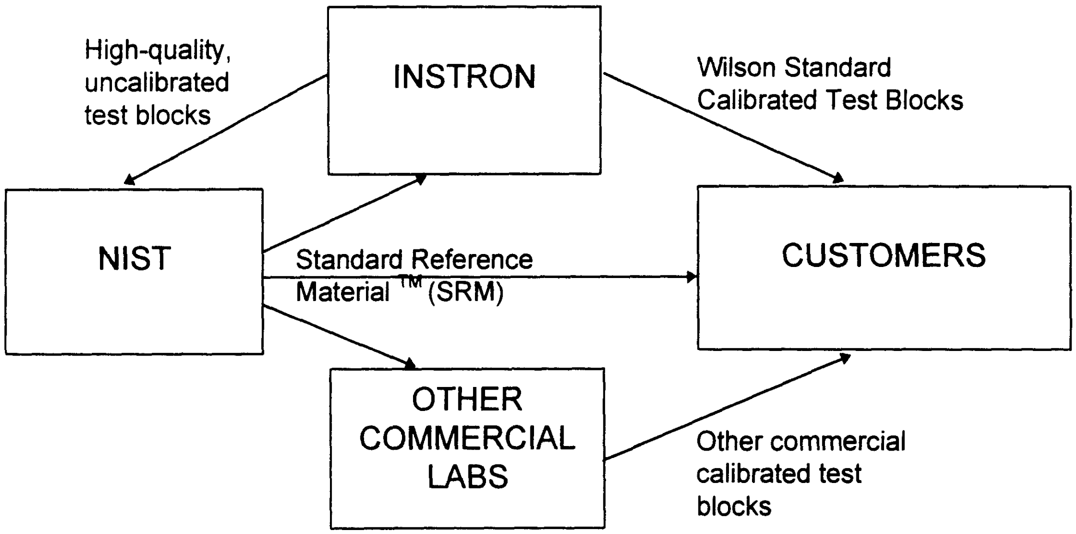
This study demonstrates how the deadweight tester measurement data supplied by NIST was used to model the sources of measurement variation. It is also cited as a crucial element to achieving statistical control of the block manufacturing process in the future [See SPC strategy, Chapter 9]. NIST currently calibrates test blocks produced by Instron to the new national standard, using the deadweight tester, to provide Standard Reference Material (SRM)™. Refer to Figure 10-2.

In order to make the SPC strategy feasible, the NIST calibrated blocks must be measured with a statistically significant quantity of indentations in order to reduce sampling error e.g $n = 15$ minimum. This quantity may extend beyond the quantity that NIST supplies to its customer base, many whose requirements are in the context of less stringent industrial applications. Hence, Instron may need to make a case to NIST for attaining blocks of higher measurement quantity.

The reduced sampling error would also yield a higher quality cross-referencing of the commercial (Instron) standard to the national standard. Accurate alignment to the new national standard will be an issue in Instron's calibration capability assessments with the general customer base. As customers trace to SRM™ blocks from NIST in setting-up their testers, there is the potential for large mean offsets ($\bar{X} - \bar{Y}$) if the Instron standards testers are not well aligned to the NIST standard.

If the current relationship with NIST cannot be maintained, Instron must consider establishing its own deadweight tester system of lower tester variation in order to effectively verify statistical control of its block manufacturing process through the SPC strategy defined. Wilson Instruments, prior to purchase by Instron Corporation, operated a deadweight Rockwell tester [13].

Figure 10-2: The relationship of NIST to commercial standardizing laboratories



Chapter 11 Conclusions and Recommendations

In order to arrive at a means to consistently control the non-uniformity of test blocks in their manufacture the influences of measurement noise of the available commercial Rockwell testers had to be addressed.

This study provided a framework to statistically characterize the systemic combination of variation sources in Rockwell hardness measurement, called the components of variances model. This model reduces the complexity of the variation problem by lumping the variation sources into two groups: block and tester. It was confirmed that Rockwell C-scale hardness measurements from either commercial or deadweight testers behave as random variables that for practical purposes can be approximated to follow a normal distribution.

The commercial testers display a measurement variation that decreases with increasing hardness (Rockwell C). As a result, statistical models and metrics must be evaluated at a prescribed hardness level. The correlation of tester variation to hardness level also adds significant complexity to the prospect of SPC control charting; control chart sets must be allocated to hardness level and part number, as well as to individual Standards testers.

The deadweight tester measurements conducted by NIST on a common test block served as a vital external leverage in estimating the relative contributions of block and commercial tester to total measurement variation. In doing so, the maximum possible standard deviation of the block was efficiently bounded by the standard deviation measured by the deadweight tester (of low measurement variation).

The variation of the commercial tester technologies was thereby found to be significantly greater than the inherent variation of a Large grade test block. A tester contribution fraction $s_{tester}/s_{measure}$ greater than 50% for the commrecial standardizing testers was determined due to the relatively low variation of the subject product.. This result compares to industry rules-of-thumb for contribution fractions of less than 10% for suitability of process control measurement. Thus, a unique (statistical) approach is required to counter the tester measurement noise in the control of the test block process

Statistical methods for ensuring that the block production process is both in *statistical control* and *capable* with respect to customer specifications are tailored to the unique constraints of the Instron test block production process. Although the majority of the data for this study was derived from the measurement behavior using Large grade blocks, the general conclusions and methods can be extended to Regular grade blocks[18].

Statistical capability metrics were developed to measure product and process performance: Calibration capability (Cc), Process capability (Cpk)[19] and Range capability (C_R). These indices

---

[18] Confirmation experiments of the statistical methodology for Regular grade blocks is recommended.
[19] Cpk is an existing, well-known capability metric that is employed to measure process accuracy for nominal hardness and was adapted to the block production environment [14].

are deemed superior to the conventional methods of direct comparison against specification ranges and tolerances in that they incorporate the probability of non-conformance over the entire measurement space. Current performance measurement of the reference standard also does not account for increased sampling error as a function of decreasing measurement sample size (e.g. for $n \leq 6$).

Calibration capability, Cc[20], measures the product performance of the global reference standard system produced by Instron. Its strength lies in that it relates the manufacturer's and customer's contribution to the overall uncertainty in the reference standard. As a result, Instron can determine the best means of satisfying customer needs by trade-off decisions of the parameters contained in the metric. The leverages for achieving capability extend beyond the reduction of block variation; opportunity lies in increasing measurement quantities, reducing tester variation and avoiding initial mean offsets. The Cc index thus serves as a guide as to *where* and *how much* to improve with an estimate of the expected result to be seen by the customer.

The Cc index simulations of actual measurement data support the following noteworthy conclusions:

- Increasing the number of measurements on the local system with the lowest total measurement variance is a low leverage policy e.g. increasing the Standards lab calibration indent quantity from 6 to say 10. Instead, more capability is marginally gained if the customer/user were to increase their measurement quantity[21],assuming the customer tester system exhibits a larger variation contribution.
- Further reduction in the nominal level of variation of the Large grade blocks of this study is an ineffective policy toward improving the customer utility of the Instron reference standard. Instead, improvement focus on the variation of the Standard tester systems, including indenter and operator influences, represents a higher Cc return policy.
- Significant Cc improvement opportunities in reducing non-uniformity of Regular-grade blocks exist.

Process capability, Cpk[22], in the context of test block manufacture measures the process accuracy in achieving the desired nominal hardness levels given as a process input. It is shown through a process control perspective that Cpk is primarily aimed at the heat treater's ability to control the hardness of each block batch through an output control feedback process in tempering.

---

[20] Equation (32) :

$$Cc = \frac{\delta x - |\bar{X} - \bar{Y}|}{3 \cdot \left[\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y} \right]^{1/2}}$$

[21] The conflict of imposing the precision burden on the customer is recognized.

[22] Equation (33) :

$$C_{pk} = \min \left\{ \frac{USL - \bar{X}}{3 \cdot \frac{s}{n^{1/2}}}, \frac{\bar{X} - LSL}{3 \cdot \frac{s}{n^{1/2}}} \right\}$$

Range capability, C_R [23], determines if the total expected measurement variation is contained within the range specification most commonly found in hardness standards such as ASTM E-18. This capability formulation was shown to be consistent with current range-based quality levels using the predictable R/s behavior found for Rockwell measurement of many blocks using a Instron 600 tester. Refer to Figure 4-6. It was determined that satisfying the range capability criteria at HRC 60 and above represents a challenge due to the tighter specification (half-) tolerances of ASTM E-18 at the higher hardness levels.

If low sample sizes ($n < 15$) are used for the capability metrics, it is suggested to compensate for the poor estimates of averages $\overline{X}$ and standard deviations s by increasing the critical capability value greater than 1.0; a criteria value of 1.5 is recommended. Otherwise, multiple replications of the capability metrics are warranted in order to minimize sampling error. In the case where the customer tester variation is determined to be significantly larger than that of the parent tester system in the Standards Laboratory, an increase in the critical value greater than one for Cpk and C_R is also suggested; such an adjustment would be made in order to avoid negative (though flawed) customer perception regarding test block uniformity [24].

A good practice for using the capability metrics is to apply the average statistic values taken from an $\overline{X}$ or s control charts of sufficient process history, as these values can be assumed typical for the process despite the small subgroup measurement size. For capability metrics to be representative of the process, statistical control must be ensured a priori.

The single biggest challenge in controlling the non-uniformity of test blocks using SPC methods is the measurement noise of the tester variation. Due to the large relative contribution of tester variation for both Large and Regular test blocks, measurement noise masks the fluctuations attributed to the block production process. As a result, the state of control for the block process cannot be directly determined using the current commercial tester technologies available in the Standards laboratory.

However, existing enablers support the conclusion that statistical process control (SPC) using conventional Shewhart control charts for X, s and R is both feasible and beneficial. The total goal of SPC when applied to the test block process is predicated not only on detecting, identifying and eliminating special causes of variation in the block process, but also in the Standards lab testers used to measure them. Once again, a lower variation tester system such as a deadweight tester must be leveraged in order to ascertain if the special cause is attributable to the tester or the block process.

A detailed SPC strategy is outlined by the author in Chapter 10 that addresses practical constraints of detail complexity. Several key aspects of this strategy stand out:

---

[23] Equation (34) : $C_R = \dfrac{Rspec}{4\sigma}$

[24] Another alternative was shown to be a reduction of internal specification requirements.

- Initial process optimization with designed experiments using deadweight tester measurement data for arriving at best-estimates of block variation.
- Control charting and root-cause feedback of a focus group of high-usage part numbers.
- Linking control charts to individual testers and to individual hardness scale levels and part numbers.
- *Copy Exactly* of process parameters and procedures for non-controlled part numbers.
- Computer automation for data acquisition and control charting.
- Bar-coding and computerized process histories to facilitate traceability and root-cause analysis.

It is also suggested to place specification limits directly on the control charts for use as real-time inspection reports on product and process performance, since (in this non-typical instance) rational control chart subgroups represent individual blocks.

The assumption should be noted that continuous operation of a deadweight tester for process measurement is cost prohibitive and impractical. In addition, a financial cost/benefit analysis is deemed necessary with regard to the investments and returns for the proposed SPC program.

SPC implementation requires investment in infrastructure, automation and personnel training. Management involvement, using the capability metrics to measure and drive improvements, is concluded to be critical in fostering statistical problem-solving and data-driven decision making. Process improvement is only gained by management focus on the corrective feedback process for eliminating root-causes of the out-of-control signals detected. Communication in problem-solving between the different parties and locations in the block manufacturing chain is regarded as a substantial hurdle to effective SPC.

There also exist many positive externalities to the methods and systems of SPC, such as process and product learning for future improvements in technology leadership, product quality and cost. In addition, the data management infrastructure can be used for enhanced customer service, to avoid errors in calibration certificates and for inventory control purposes.

Significant discipline is required in maintaining control and consistency of standardized process procedures across all test block types in order to reap the improvement benefits of SPC-driven root-cause analysis. It is also the essence of the 'copy exactly' strategy. Such task standardization of individual operations will require adjustment in designing the work roles of production associates who may feel their autonomy and variety in defining work methods is threatened. The managerial approach advocated by the author is to increase the involvement of production associates in task design and improvement problem-solving.

The Calibration Capability index empowers the manufacturer and customer to make better decisions on the quality and performance of their commercial reference standard system for Rockwell hardness testing. Comprehensive SPC will allow Instron to perpetually improve in order to continue in exceeding customer expectations relative to the competition. The keys to improvement lie in statistical modeling and the low variation of a secondary deadweight tester. A continued relationship with NIST is thus deemed important to fulfill the latter ingredient.

'Through discipline lies freedom' [25]

11.1 Recommendations for Further Investigation

As this study focused only on the Rockwell 'C'-scale, validation of the statistical models on other hardness levels is recommended. Development of statistical regression models in order to predict the level of measurement variation as a function of hardness would be beneficial in understanding the response behavior of the commercial Rockwell testers.

It is recommended that the predictable R/s behavior (See Figure 4-6) at the small measurement subgroup sizes of n = 5 or 6 should be confirmed with more block samples or SPC data, such that the R-chart can be eliminated. The R-chart's purpose is only to directly ensure specification conformance; the s-chart suffices to control and characterize block and measurement variation. The condensed R/s behavior would allow the R-chart to be replaced by an equivalent specification limit on the s-chart e.g. s_{USL}= Range Spec/4.

Since Regular blocks were not measured by a deadweight tester in this study, the relative contribution of block non-uniformity is estimated only by the results of the measurement variation from the commercial testers. If the Regular grade blocks are maintained as a product offering, it is recommended to test Regular blocks over the full hardness spectrum using a deadweight tester. The components of variances model can thus be applied as demonstrated in this study.

Application of the Calibration Capability index to characterize a representative sampling of field tester systems in customer environments is further recommended. Analysis of the input parameters to the index will help Instron determine where the leverages for global system improvement in the commercial reference standard lie.

A financial cost/benefit analysis for the elements of the SPC strategy proposed in this study is necessary for managerial decision-making. In order to do so, some formal layout of the information system (data acquisition, control charting and data management) is required. Finally it is recommended to test the communication ability of the current production system for the critical root-cause analysis and problem-solving activities of SPC. Co-location of output measurement and key process operations may be called for in order to enable a rapid feedback process.

[25] Author unknown.

LITERATURE REFERENCES

[1] Beckwith, Thomas G., Lienhard, John H., Marangon, Roy D., *Mechanical Measurement*, 5th ed., Addison-Wesley, Reading, MA, 1993

[2] Cieplak, Joseph J., Tobolski, Edward L., Williams, Dennis M., "Gage Repeatablility and Reproducability Studies of Rockwell Scale Hardness Testers", *Factors That Affect the Precision of Mechanical Tests, ASTM STP 1025*, American Society of Metals, Philadelphia, 1989, pp. 52-60

[3] "Hardness Test Blocks and Indentors", OIML, Bureau International de Metrologie Legale, Paris, France, 1984

[4] ASTM E-18, "Standard Test Methods for Rockwell Hardness and Rockwell Superficial Hardness of Metallic Materials", *Annual Book of ASTM Standards*, ASTM, Feb. 1993, pp. 118-130.

[5] "Fundamentals of Rockwell Hardness Testing", Wilson Instruments, Division of Instron Corporation, Canton, MA, 1988

[6] DIN 51303, "Haertevergleichsplatten fuer die Pruefung statisher Haertepruefgeraete" (Standardized blocks for the verification of static hardness machines), Deutsches Institut fuer Normung e.V., November, 1983

[7] DIN 51304, "Pruefung von Haertepruefgeraeten mit Eindringtiefen-Messeinrichtung" (Testing of hardness testing machines with a depth-measuring device), Deutsches Institut fuer Normung e.V., September, 1983

[8] JIS B 7730, "Standardized Blocks of Rockwell and Rockwell Superficial Hardness", Japanese Standards Association, 1980

[9] NIS 0406, "Accreditation for the Calibration and Verification of Hardness Testing Equipment", 2nd ed., NAMAS Executive, National Physical Laboratory, Teddington, Middlesex, United Kingdom, April 1993

[10] Atwood, Corwin L., Engelhardt, Max, "Techniques for Uncertainty Analysis of Complex Measurement Processes", Journal for Quality Technology, Vol. 28, No. 1, The American Society for Quality Control, January, 1996. pp. 1-11

[11] Ramsey, Patricia, Ramsey, Philip, "Simple Tests for Normality of Small Samples", Journal for Quality Technology, Vol. 22, No. 4, The American Society for Quality Control, October 1990. pp. 299-308

[12] Delcourt, Michelle L., Lohmann, Andrea D., Stiller, David L., *An Investigation of the Variability in the Performance of Rockwell Hardness Testers Produced by the Wilson Instruments Division, Instron Corporation, Canton Massachusetts*, Master's Project Report, Operations Research and Industrial Engineering, Cornell University, May 1993

[13] DeBellis, Anthony, "Maintaining Rockwell Hardness Standards", *Materials Research and Standards, Vol. 12, No. 8,* American Society for Testing and Materials, Aug. 1972, pp.28-30

[14] DeVor, Richard E., Chang, Tsong-how, Sutherland, John W., *Statistical Quality Design and Control, Contemporary Concepts and Methods*, Macmillan, New York, NY, 1992

[15] Hogg, Robert V., Ledolter, Johannes, *Applied Statistics for Engineers and Physical Scientists*, 2nd ed., Macmillan, New York, NY, 1992

[16] Duncan, Acheson J., *Quality Control and Industrial Statistics*, 4th ed., Irwin, Homewood, IL, 1974

[17] Dougherty, Edward R., *Probability and Statisitics for the Engineering, Computing and Physical Sciences*, Prentice-Hall, Englewood Cliffs, NJ, 1990

[18] Kiemele Mark J., Schmidt, Stephen R., *Basic Statistics, Tools for Continuous Improvement*, 3rd ed., Air Academy Press, Colorado Springs, CO, 1993

[19] Hardt, David E., "Manufacturing Processes and Process Control", Laboratory for Manufacturing and Productivity, Massachusetts Institute of Technology, Feb. 1995

[20] Thelning, Karl-Erik, *Steel and its Heat Treatment*, 2nd ed., Butterworths, United Kingdom, 1984

[21] Phadke, Madhav S., *Quality Engineering using Robust Design,* Prentice Hall, Englewood Cliff, NJ, 1989

[22] Yamamoto, H. , Yamamoto, I.,"The Development of Standard Hardness Blocks", *VDI Berichte*, 1985, pp. 281-294

[23] Nakai, Isao, Inoue, Katsuhiko, Ishida, Hajime, Yano, Hiroshi, "Development of Heat Treatment Technology for the Standard Hardness Blocks", *ASI Symposium on Taguchi Methods*, Detroit, Nov. 1991

[24] Shiba, Shoji, *A New American TQM,* Center for Quality Management, Productivity Press, Portland, OR, 1993

[25] Shecter, Edwin S., *Managing world-class quality: A primer for executive and managers*, Marcel Dekker/ASQC Quality Press, Milwaukee, WI, 1992, Ch. 2

[26] "This Guide has Software You Need; Quality Software Selection Guide", *Quality*, Vol. 34, Dec., 1995, pp. 26-46

[27] "Automation means more data can be collected and analyzed faster with less chance of error", *Chemical Engineering*, Sept. 1995

[28] Teresko, John, "Rockwell International, El Paso, Texas", *Industry Week*, October 16, 1995

[29] Scott, John, Golkin, Ken, "Rapid deployment of automated SPC systems; statistical process control in the aerospace manufacturing industry", *Industrial Engineering*, Aug. 1993

[30] Blunt, Bryan T., "When should control charts be implemented? Statistical process control and process capability", *Modern Machine Shop*, April 1993

[31] Kline, Mathew, "Short run environment tests the limits of traditional SPC", *Paperboard Packaging*, March 1993

[32] Bajaria, Hans J., "Is your investment paying off? Manufacturers to evaluate statistical process control programs", *Industrial Finishing*, April, 1989

[33] Hare, Lynne B., Hoerl, Roger W., Hromi, John D., Snee, Ronald D., "The Role of Statistical Thinking in Management", *Quality Progress*, February, 1995

[34] Barnett, Andrew J., Andrews, Richard W., "Are You Getting the Most Out Of Your Control Charts ?", *Quality Progress*, November 1994

[35] Heifetz, Ronald A., *Leadership without Easy Answers*, Harvard University Press, Cambridge, MA, 1994

[36] Drake, Alvin W., Keeney, Ralph L., *Decision Analysis*, No. 37-2100, MIT Center for Advanced Engineering Study, Cambridge, MA, 1978

[37] Austin, Amy B., *Management and Scheduling Aspects of Increasing Flexibility in Manufacturing*, Master's Thesis, MIT Sloan School of Management, Cambridge, MA, 1993

[38] Spiegel, Murray R., *Schaum's Outline of Theory and Problems of Probability and Statistics*, McGraw and Hill, 1994

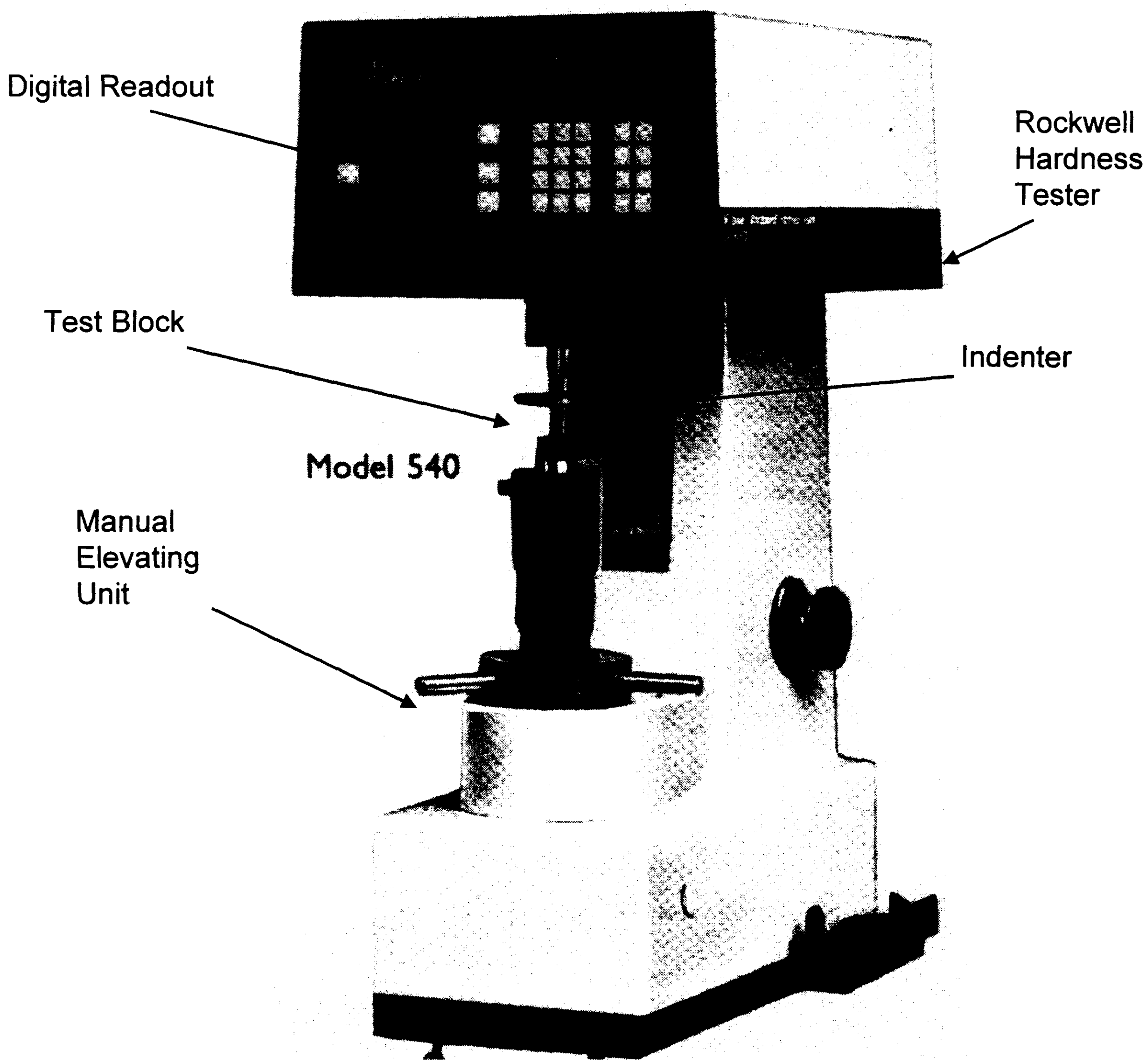
[39] Klein, Janice A., "A Reexamination of Autonomy in Light of New Manufacturing Practices", Human Relations, Vol. 44, Tavistod Institute for Human Relations, London, U.K., 1991

[40] Mlynarczyk, Michelle, *Improving Synergy in Multi-Site Microprocessor Manufacturing: An Analysis of the Copy Exactly Approach*, Master's Thesis, MIT Sloan School of Management, Cambridge, MA, 1995

APPENDICES

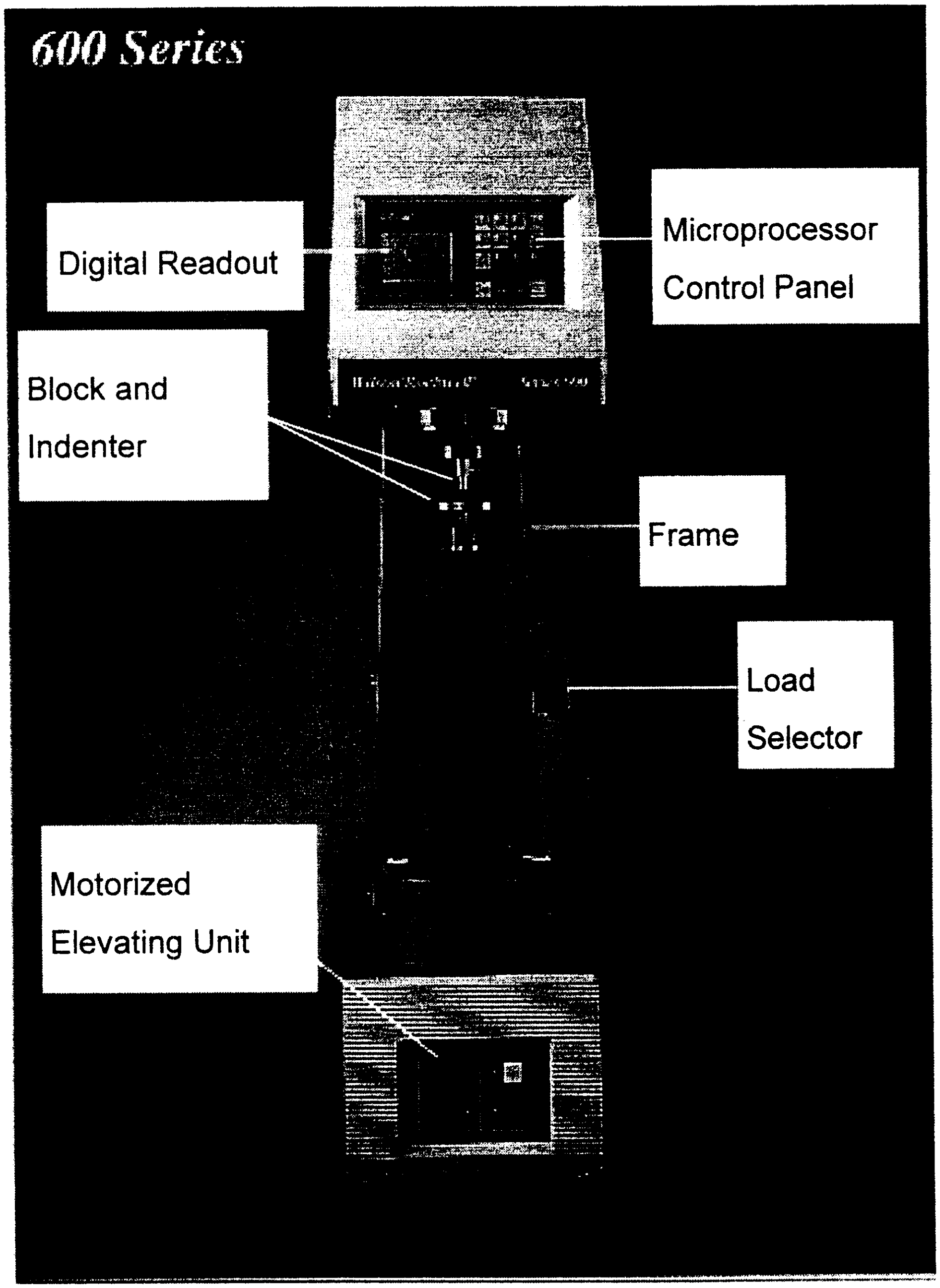
APPENDIX A Equipment Diagrams of Instron Rockwell© Tester Systems

Figure A-1: 500 Series Standards Tester



Ref.: Series 500 Rockwell Hardness Testers, Catalog 1817, Wilson Instruments, Division of Instron Corporation

Figure A-2: 600 Series Instron Tester



APPENDIX B Diagrams of Instron Standard Rockwell© Test Blocks

Figure B-1: Large Grade Test Block

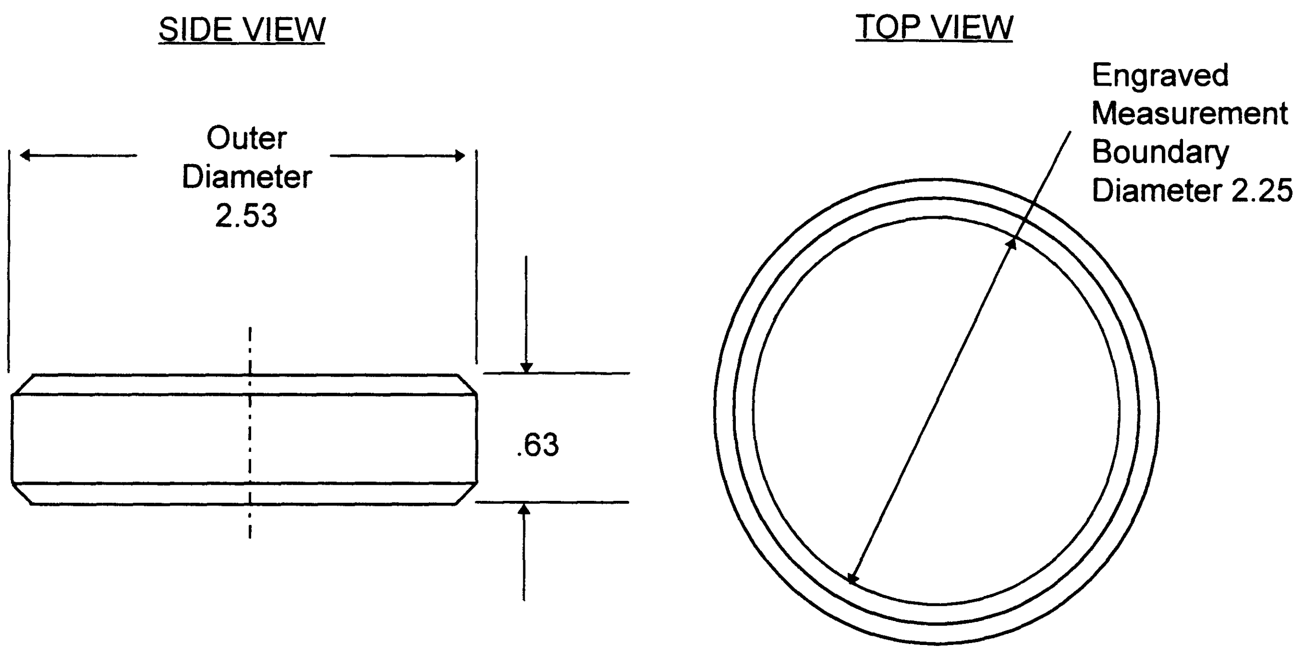
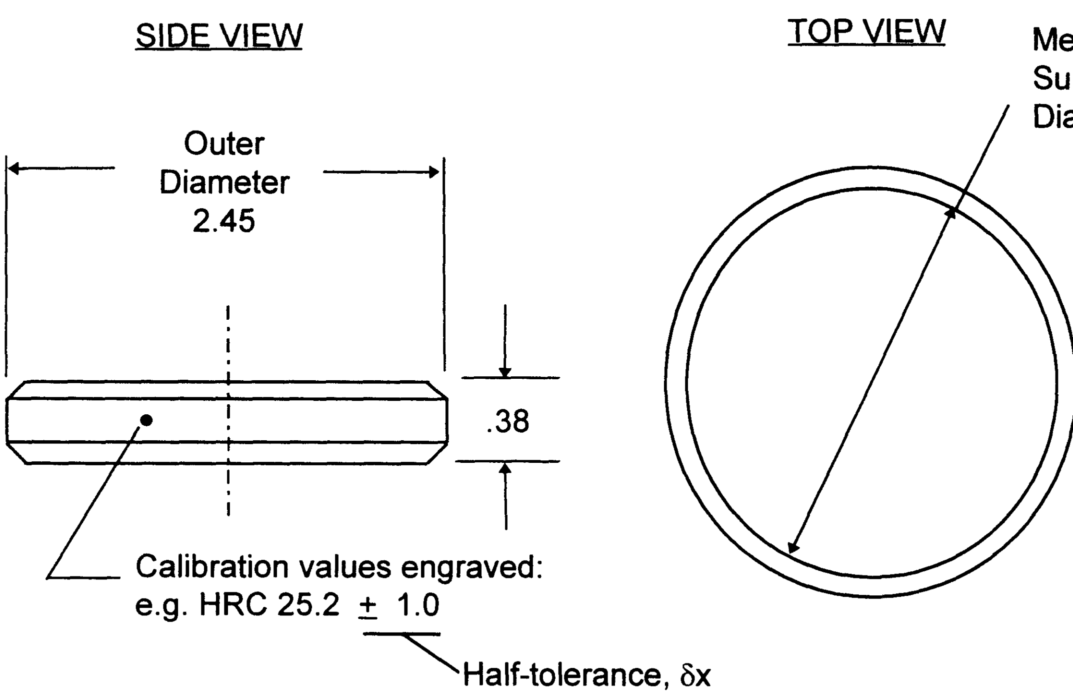


Figure B-2: Regular Grade Test Block



Note: All dimensions in inches.

APPENDIX C **Test Block Rockwell Measurement Data**

Table C-1: NIST Measurement Data using a Deadweight Tester

Measure No.	Block Ser. No. 95I30005	Block Ser. No. 95I40004	Block Ser. No. 95I50005	Block Ser. No. 95I60001
	HRC	HRC	HRC	HRC
1	29.96416	40.49616	49.94792	60.77507
2	29.97702	40.60641	49.87524	60.76864
3	29.93499	40.54659	49.93753	60.75826
4	30.08232	40.54857	49.87821	60.75529
5	29.98641	40.57329	49.95237	60.76518
6	30.02843	40.53324	49.85201	60.77012
7	29.99877	40.49764	49.93655	60.77902
8	30.02547	40.53967	49.85942	60.76024
9	30.09419	40.46254	49.93902	60.73947
10	30.09270	40.51989	49.94001	60.77507
11	29.99976	40.51989	49.99736	60.71080
12	29.98443	40.49517	49.88117	60.76221
13	29.96021	40.40915	49.94248	60.75331
14	30.01558	40.58911	49.88266	60.74540
15	29.91522	40.49023	49.86881	60.73650
16	30.07342	40.47045	49.92171	60.77606
17	30.02349	40.46155	49.96028	60.75035
18	30.12286	40.44623	49.91430	60.75529
19	29.94933	40.53324	49.88958	60.76370
20	29.96861	40.46551	49.87228	60.75529
21	29.98789	40.35922	49.92468	60.75480
22	30.01261	40.50012	49.84904	60.74095
23	29.99432	40.39283	49.89452	60.76815
24	30.02893	40.36663	49.91034	60.78149
25	29.98938	40.47985	49.88414	60.75035
26	30.00915	40.41063	49.84212	60.75183
27	29.97553	40.48331	49.86634	60.75974
28	30.08875	40.48973	49.83174	60.74046
29	29.90533	40.51940	49.85942	60.72563
30	30.08430	40.43980	49.93012	60.74540
31	29.97257	40.51445	49.88117	60.73749
32	29.91176	40.43436	49.91875	60.73700
33	29.97801	40.40124	49.89897	60.77012
34	30.05068	40.48973	49.88266	60.73057
35	30.04425	40.42349	49.89353	60.71574
36	29.94983	40.44722	49.95731	60.72217
37	30.00668	40.54461	49.87277	60.75529
38	29.97059	40.46007	49.86783	60.75035
39	30.13671	40.51940	49.88958	60.76024
40	29.92115	40.45216	49.85794	60.76765
41	29.92362	40.45414	49.89848	60.75974
42	29.99580	40.44277	49.89057	60.75875

APPENDIX C

Table C-1: *Continued*

Measure No.	Block Ser. No. 95I30005	Block Ser. No. 95I40004	Block Ser. No. 95I50005	Block Ser. No. 95I60001
43	29.95526	40.51099	49.92864	60.77902
44	30.06650	40.41409	49.89601	60.78100
45	29.97108	40.45760	49.88216	60.73799
46	30.12088	40.41805	49.87376	60.75134
47	30.11396	40.56439	49.90837	60.74194
48	30.03338	40.50704	49.96621	60.77704
49	29.93499	40.53621	49.90194	60.73799
50	30.07787	40.51149	49.85893	60.74046
51	30.08183	40.55450	49.88908	60.78940
52	29.95526	40.46798	49.91825	60.80325
53	29.98938	40.39976	49.89798	60.72612
54	29.98295	40.46749	49.86881	60.78495
55	29.90039	40.43288	49.89353	60.76073
56	29.92461	40.49863	49.93061	60.75826
57	29.82969	40.38740	49.92913	60.79484
58	30.04030	40.46452	49.91578	60.72859
59	29.84996	40.45859	49.88464	60.81165
60	30.01904	40.55747	49.91133	60.75035
61	30.04376	40.30088	49.88760	60.77210
62	29.93005	40.29742	49.88365	60.73848
63	29.92610	40.30928	49.87870	60.77507
64	30.08133	40.36564	49.88859	60.73700
65	30.08529	40.49666	49.88859	60.76122
66	29.89050	40.45760	49.88859	60.76765
67	29.96268	40.38492	49.86881	60.77210
68	29.98839	40.37059	49.88464	60.72662
69	--	40.38492	49.84953	60.75084
70	--	40.30681	49.85695	60.74145
71	--	40.44277	49.85794	60.75925
72	--	40.44475	49.90639	60.76370
73	--	40.38888	49.88908	60.78693
74	--	40.47144	49.89699	60.79089
76	--	40.39580	49.86733	60.75826
77	--	40.48034	--	--

Summary Statistics:

Average	29.99839	40.46234	49.89599	60.75728
Max	30.13671	40.60641	49.99736	60.81165
Min	29.82969	40.29742	49.83174	60.71080
Range	0.3070	0.3090	0.1656	0.1009
Std.Dev.	0.0673	0.0680	0.0325	0.0202

Note: Refer to the measurement pattern of Figure E-1, Appendix E

APPENDIX C Test Block Measurement Data

Table C-2: 600S Measurements of Large Grade Blocks

Date	1/26/96	1/26/96	1/26/96	1/26/96
Tester	A653R MT-4	A653R MT-4	A653R MT-4	A653R MT-4
Serial No.	97328502	97328502	97328502	97328502
Indentor SN	95621105	95621105	95621105	95621105
Operator	HJL	HJL	HJL	HJL
Block SN	95130005	95140004	95150005	95160001
Measure No.	HRC	HRC	HRC	HRC
1	30.10	40.49	49.88	60.32
2	30.62	40.62	49.89	60.41
3	30.30	40.41	49.90	60.41
4	30.08	40.61	49.79	60.58
5	30.22	40.45	49.95	60.49
6	30.07	40.43	49.99	60.40
7	30.20	40.45	50.04	60.53
8	30.26	40.61	49.87	60.38
9	30.66	40.55	49.96	60.38
10	30.42	40.48	49.94	60.38
11	30.40	40.51	49.90	60.38
12	30.49	40.44	49.96	60.38
13	30.22	40.60	49.89	60.43
14	30.47	40.34	49.96	60.43
15	30.18	40.48	49.81	60.26
16	30.25	40.53	50.00	60.36
17	30.47	40.41	49.92	60.38
18	30.45	40.65	49.94	60.41
19	30.24	40.63	49.89	60.45
20	30.36	40.63	49.96	60.49
21	30.18	40.46	50.01	60.41
22	30.07	40.55	49.88	60.30
23	30.47	40.52	50.00	60.34
24	30.29	40.53	49.98	60.35
25	30.31	40.45	49.96	60.32
26	30.20	40.64	49.89	60.39
27	30.11	40.31	49.97	60.43
28	30.41	40.61	50.02	60.52
29	30.11	40.59	50.01	60.23
30	30.32	40.48	49.92	60.44

Summary Statistics:

Average	30.30	40.52	49.94	60.40
Std. Dev.	0.16	0.09	0.06	0.08
Max.	30.66	40.65	50.04	60.58
Min.	30.07	40.31	49.79	60.23
Range	0.59	0.34	0.25	0.35
25 Pctle.	30.19	40.45	49.89	60.37
75 Pctle.	30.42	40.61	49.98	60.43

Note: Randomized measurement locations

APPENDIX C Test Block Measurement Data

Table C-3: 500S Measurements of Large Grade Blocks

Date	1/26/96	1/26/96	1/26/96	1/26/96
Tester	B523R	B523R	B523R	B523R
Serial No.	80195408	80195408	80195408	80195408
Indentor SN	9401911	9401911	9401911	9401911
Operator	HJL	HJL	HJL	HJL
Block SN	95130005	95140004	95150005	95160001
Measure No.	HRC	HRC	HRC	HRC
1	30.1	40.3	49.3	60.2
2	30.4	40.2	49.5	59.8
3	30.3	40.3	49.4	60.2
4	30.4	40.5	49.7	60.0
5	30.1	40.4	49.7	60.1
6	30.3	40.2	49.6	60.0
7	30.0	40.5	49.5	60.0
8	30.3	40.4	49.8	60.5
9	30.1	40.2	49.6	60.1
10	30.1	40.3	49.6	60.0
11	30.7	40.4	49.6	60.0
12	30.3	40.6	49.6	60.0
13	30.3	40.5	49.5	60.1
14	30.4	40.3	49.7	60.1
15	30.2	40.4	49.7	60.2
16	30.3	40.2	49.5	60.2
17	30.4	40.7	49.6	60.1
18	30.3	40.2	49.8	59.9
19	30.1	40.2	49.6	60.1
20	30.5	40.4	49.7	60.0
21	30.4	40.3	49.8	60.2
22	30.5	40.2	49.4	60.2
23	30.2	40.4	49.5	60.1
24	30.2	40.4	49.6	60.1
25	30.3	40.3	49.7	59.9
26	30.0	40.2	49.6	60.1
27	30.3	40.1	49.6	60.1
28	30.2	40.5	49.5	60.1
29	30.6	40.5	49.6	60.1
30	30.5	40.3	49.5	60.2

Average	30.3	40.3	49.6	60.1
Std. Dev.	0.2	0.1	0.1	0.1
Max.	30.7	40.7	49.8	60.5
Min.	30.0	40.1	49.3	59.8
Range	0.7	0.6	0.5	0.7
25 Percentile	30.2	40.2	49.5	60.0
75 Percentile	30.4	40.4	49.7	60.2

Note: Randomized measurement locations

Table C-4: 600S Measurements of Regular Grade Blocks

Date	1/26/96	1/26/96	1/26/96
Tester	A653R MT-4	A653R MT-4	A653R MT-4
Serial No.	97328502	97328502	97328502
Indentor SN	95621105	95621105	95621105
Operator	HJL	HJL	HJL
Block SN	H00128	G00390	R02589
Measure No.	HRC	HRC	HRC
1	26.40	45.96	63.69
2	25.78	46.13	63.61
3	26.83	46.04	63.76
4	26.21	45.80	63.77
5	26.18	46.13	63.77
6	26.20	46.20	63.71
7	26.43	46.09	63.68
8	26.31	46.12	63.65
9	25.95	46.08	63.74
10	26.30	45.80	63.66
11	26.18	46.14	63.67
12	25.96	46.08	63.70
13	26.14	46.15	63.73
14	26.26	46.16	63.77
15	26.44	45.97	63.67
16	26.05	46.02	63.76
17	26.46	46.11	63.72
18	26.38	45.73	63.70
19	26.26	46.07	63.71
20	26.26	46.14	63.70
21	25.71	45.97	63.62
22	26.40	45.74	63.49
23	26.42	46.08	63.63
24	26.42	46.13	63.67
25	26.30	45.78	63.89
26	26.04	46.15	63.80
27	26.36	46.04	63.53
28	26.21	46.16	63.74
29	25.98	46.01	63.75
30	26.23	45.97	63.72

Summary Statistics:

Average	26.24	46.03	63.70
Std. Dev.	0.22	0.14	0.08
Max.	26.83	46.20	63.89
Min.	25.71	45.73	63.49
Range	1.12	0.47	0.40
25 Percentile	26.15	45.97	63.67
75 Percentile	26.40	46.13	63.75

Note: Randomized measurement locations

Table C-5: 500S Measurements of Regular Grade Blocks

Date	1/26/96	1/26/96	1/26/96
Tester	B523R	B523R	B523R
Serial No.	80195408	80195408	80195408
Indentor SN	9401911	9401911	9401911
Operator	HJL	HJL	HJL
Block SN	H00128	G00390	R02589
Measure No.	HRC	HRC	HRC
1	25.9	45.6	63.2
2	26.0	45.6	63.6
3	26.2	45.5	63.5
4	26.3	45.9	63.7
5	26.2	45.7	63.4
6	26.9	45.9	63.6
7	26.4	45.9	63.5
8	26.2	46.0	63.5
9	26.6	45.9	63.5
10	26.7	46.1	63.5
11	26.1	46.2	63.5
12	26.4	46.1	63.5
13	26.4	46.2	63.4
14	26.3	45.9	63.2
15	26.3	46.2	63.6
16	26.6	45.9	63.6
17	26.4	45.7	63.6
18	26.5	45.9	63.5
19	26.3	46.1	63.4
20	26.4	46.0	63.5
21	26.6	45.9	63.4
22	26.4	46.1	63.6
23	26.0	45.7	63.6
24	26.4	46.1	63.6
25	26.1	46.2	63.7
26	26.9	45.6	63.4
27	26.5	46.0	63.5
28	26.8	46.0	63.7
29	26.5	45.9	63.5
30	26.1	46.1	63.4

Summary Statistics:

Average	26.4	45.9	63.5
Std. Dev.	0.3	0.2	0.1
Max.	26.9	46.2	63.7
Min.	25.9	45.5	63.2
Range	1.0	0.7	0.5
25 Percentile	26.2	45.9	63.4
75 Percentile	26.5	46.1	63.6

Note: Randomized measurement locations

APPENDIX C Test Block Measurement Data

Table C-6: 600R Measurements of Large Grade Blocks

Date	10/26/95	10/25/95	10/25/95	10/24/95
Tester	653 C	653 C	653 C	653 C
Serial No.	97331502	97331502	97331502	97331502
Indentor SN	95621105	95621105	95621105	95621105
Operator	HJL	HJL	HJL	HJL
Block SN	95125016	95145005	95145006	95163020
Measure No.	HRC	HRC	HRC	HRC
1	25.53	45.35	45.44	63.60
2	25.54	45.40	45.54	63.40
3	25.67	45.40	45.37	63.41
4	25.33	45.32	45.60	63.36
5	25.48	45.32	45.70	63.55
6	25.34	45.57	45.21	63.52
7	25.52	45.41	45.32	63.68
8	25.48	45.41	45.39	63.65
9	25.67	45.67	45.47	63.60
10	25.43	45.73	45.54	63.59
11	25.62	45.68	45.59	63.59
12	25.54	45.70	45.49	63.62
13	25.35	45.59	45.41	63.64
14	25.31	45.66	45.61	63.70
15	25.49	45.23	45.71	63.58
16	25.48	45.68	45.36	63.62
17	25.49	45.83	45.59	63.60
18	25.41	45.66	45.53	63.60
19	25.50	45.62	45.40	63.52
20	25.54	45.64	45.55	63.71
21	25.74	45.40	45.43	63.66
22	25.62	45.61	45.65	63.60
23	25.49	45.71	45.64	63.65
24	25.62	45.19	45.48	63.62
25	25.63	45.44	45.32	63.65
26	25.61	45.74	45.58	63.56
27	25.80	45.63	45.23	63.60
28	25.73	45.80	45.52	63.91
29	25.55	45.32	45.50	63.61
30	25.55	45.61	45.41	63.71
31	25.57	45.63	45.59	63.79
32	25.61	45.31	45.47	63.55
33	25.58	45.59	45.48	63.61
34	25.54	45.59	45.64	63.68
35	25.59	45.42	45.66	63.69
36	25.74	45.43	45.73	63.68
37	25.86	45.62	45.62	63.72
38	25.43	45.62	45.47	63.66
39	25.67	45.67	45.59	63.65
40	25.63	45.71	45.46	63.57
41	25.78	45.69	45.48	63.63
42	25.53	45.51	45.52	63.59
43	25.67	45.47	45.64	63.70
44	25.44	45.67	45.62	63.76
45	25.56	45.62	45.64	63.57
46	25.88	45.49	45.59	63.66
47	25.56	45.55	45.46	63.64
48	25.72	45.57	45.41	63.75
49	25.62	45.45	45.27	63.68
50	25.58	45.48	45.52	63.64
51	25.81	45.57	45.57	63.60
52	25.56	45.27	45.62	63.59
53	25.40	45.69	45.58	63.69
54	25.40	45.34	45.32	63.76
55	25.71	45.33	45.64	63.76
56	25.42	45.79	45.47	63.58
57	25.53	45.35	45.67	63.79
58	25.72	45.46	45.53	63.65
59	25.62	45.67	45.55	63.51
60	25.73	45.65	45.67	63.66

Summary Statistics:

Average	25.58	45.54	45.52	63.63
Std. Dev.	0.131	0.156	0.121	0.092
Max.	25.88	45.83	45.73	63.91
Min.	25.31	45.19	45.21	63.36
Range	0.57	0.64	0.52	0.55
25 Pctle.	25.49	45.41	45.455	63.59
75 Pctle.	25.67	45.67	45.6125	63.68

Note: Randomized measurement locations

APPENDIX D Sample Calibration Certificate

Wilson® Rockwell®
HARDNESS TEST BLOCK
X STANDARD ___AA CLASS

X̄: 45.5 HR C
Average Scale
R: 0.2
Uniformity

CERTIFICATE of CALIBRATION

The Wilson Standards Laboratory certifies that the calibration results recorded on this certificate are true and correct, and that this test block has been manufactured and standardized in accordance with ASTM standard E 18.

This test block was calibrated on a verified laboratory standard tester with load and depth measuring devices traceable to NIST (NBS)*. The Wilson calibration program conforms to MIL-STD-45662A. The hardness standards maintained by the Wilson Standards Laboratory are those originally developed by Stanley P. Rockwell in 1919.

SERIAL NO. R01234

*LAB NO. 822.077256035

*LAB NO. 821/252017/93

Wilson® Instruments
Division of Instron Corporation
100 Royall Street, Canton, MA 02021
(617) 575-6000 FAX: (617) 575-5770

CERTIFIED CALIBRATION RESULTS

1. 45.6
2. 45.4
3. 45.6
4. 45.4
5. 45.5
6. 45.4

Use this standardized hardness test block to verify the operation of a tester in accordance with ASTM E18. A verified tester meets both Repeatability and Error criteria set forth in the standard.

As established by ASTM E18. the tolerance to be used for Error calculation at this hardness value is: ± 1.0.

In use, this block should be supported on a pedestal spot anvil. Tests must not be made any closer together than 3 diameters since previous indentations distort the hardness of the surrounding area. Only the surface with the Wilson Standards Laboratory mark is standardized. No other surface should be used. The standardized surface must not be reworked.

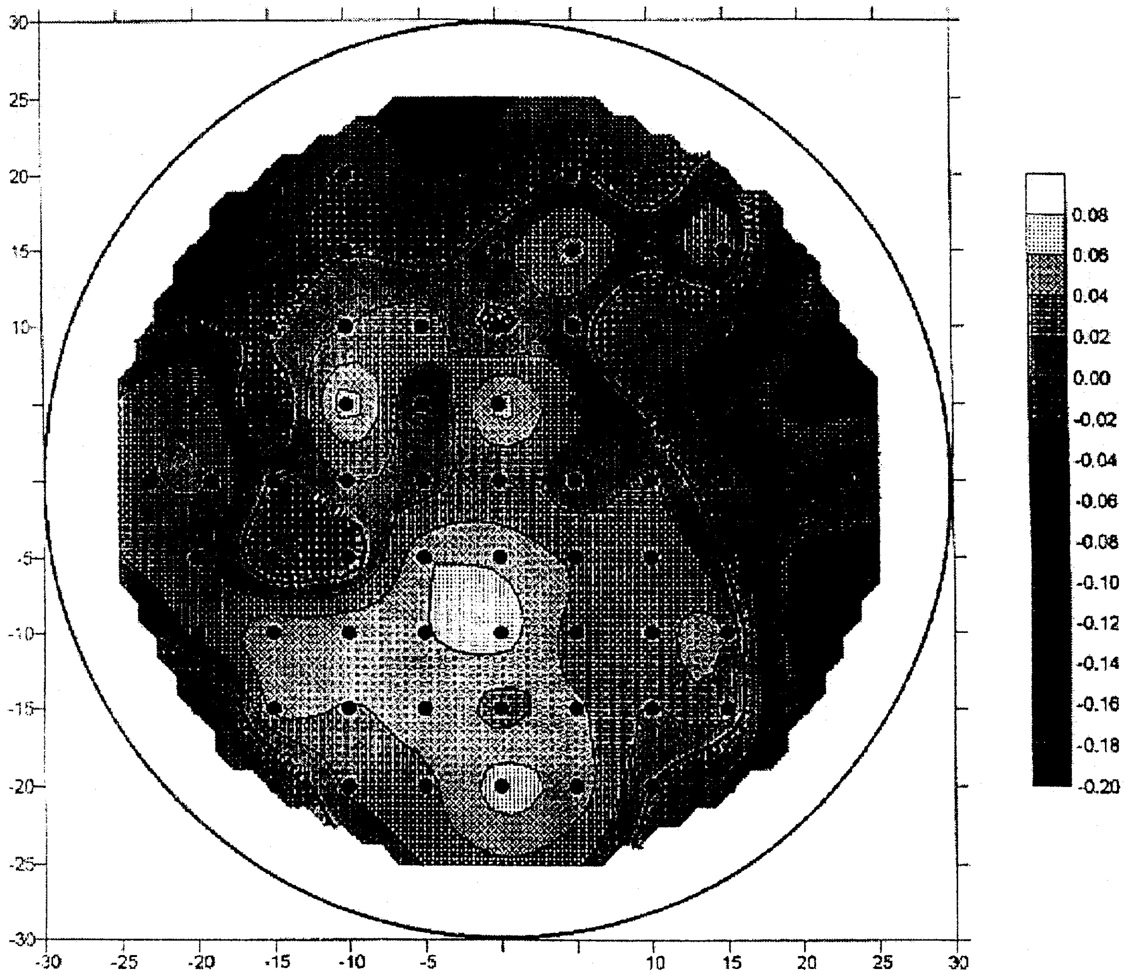
DATE 10.23.95

INSPECTOR 472

Wilson®, Rockwell® and Brale® are registered trademarks of Wilson Instruments, Inc.
Made in USA

APPENDIX E Test Block Measurement Patterns

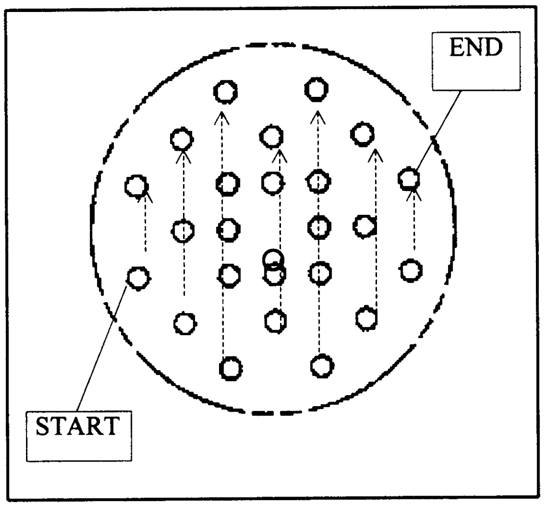
Figure E-1: Typical NIST Sequential Measurement Pattern (68 to 77 indents)



Ref.: Courtesy of the National Institute of Standards and Technology, Gathersburg, MD

APPENDIX E Test Block Measurement Patterns

Figure E-2: Instron Measurement Pattern (25 indents)



APPENDIX F   Q-Q Plots for Normality Goodness-of-Fit

Figure F-1:

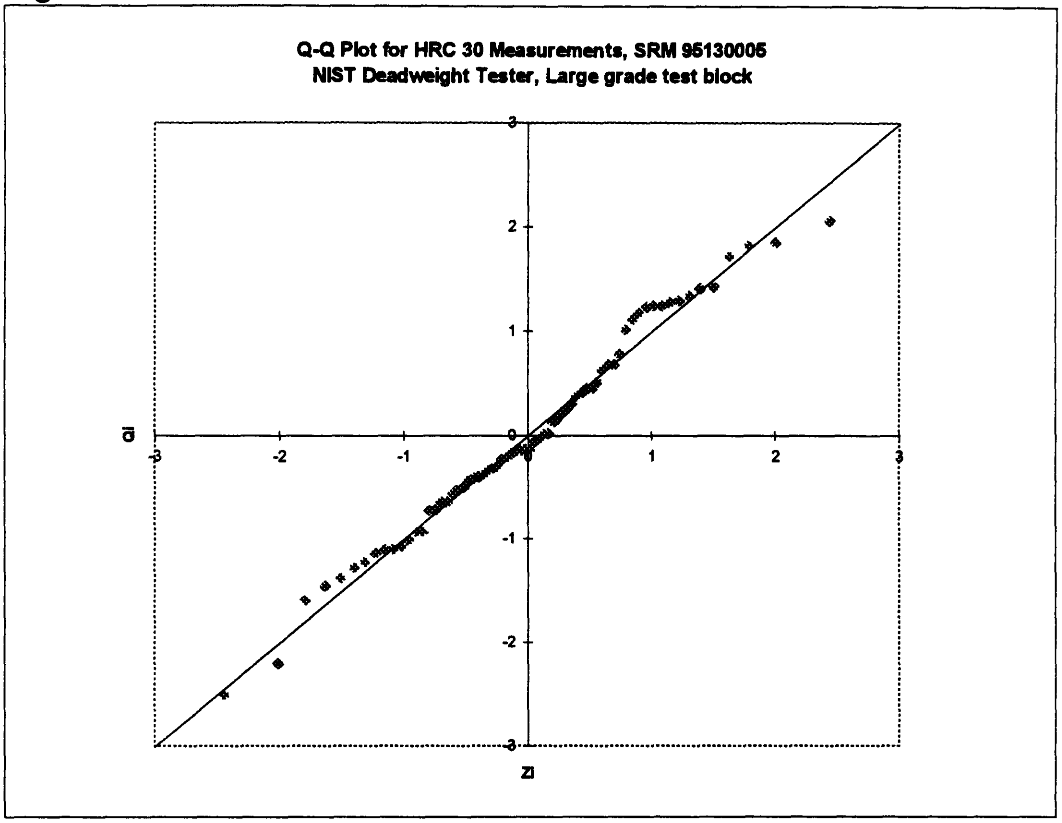


Figure F-2:

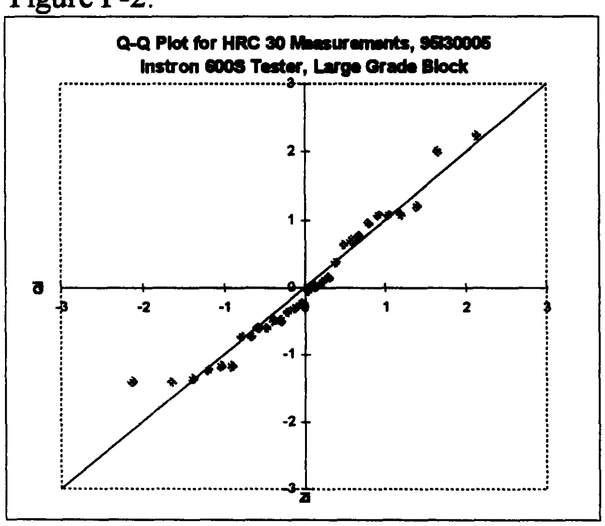
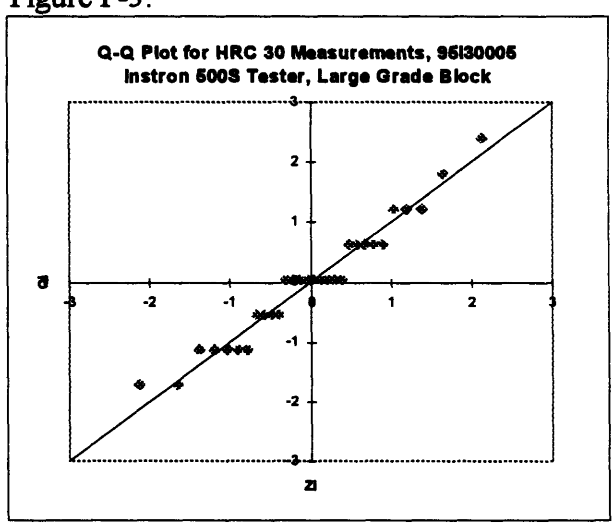


Figure F-3:



APPENDIX F <u>**Q-Q Plots for Normality Goodness-of-Fit**</u>

Figure F-4:

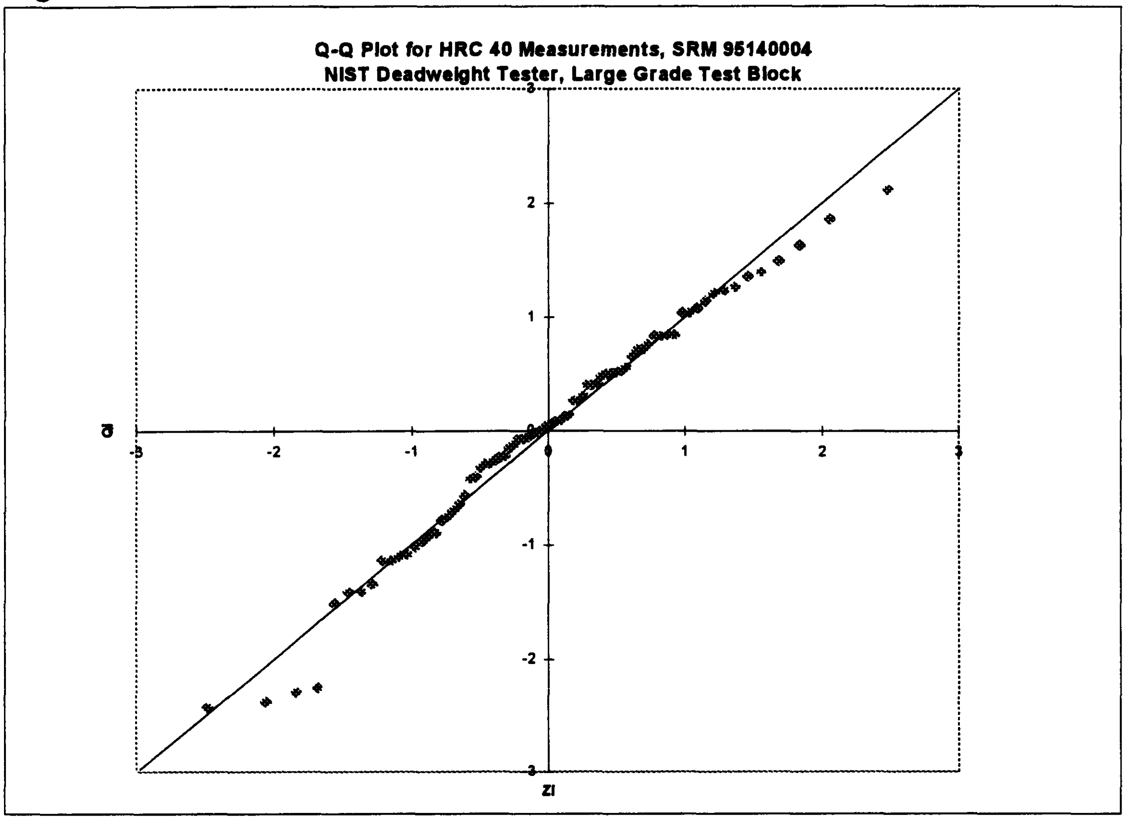


Figure F-5:

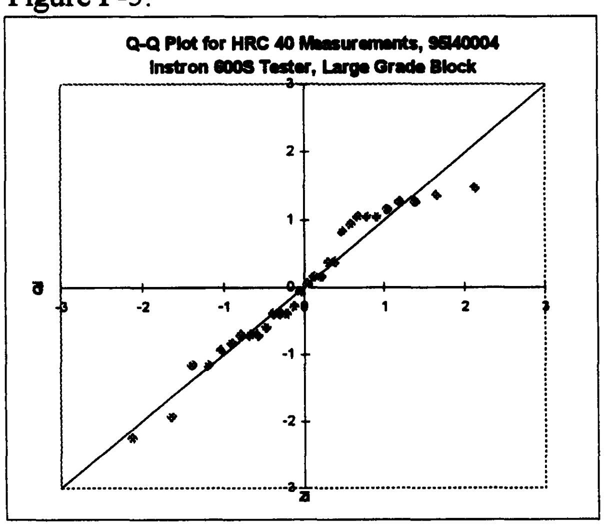
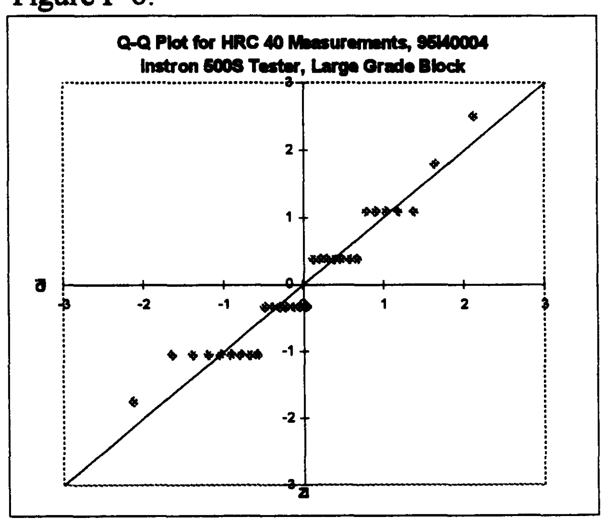


Figure F-6:



APPENDIX F **Q-Q Plots for Normality Goodness-of-Fit**

Figure F-7:

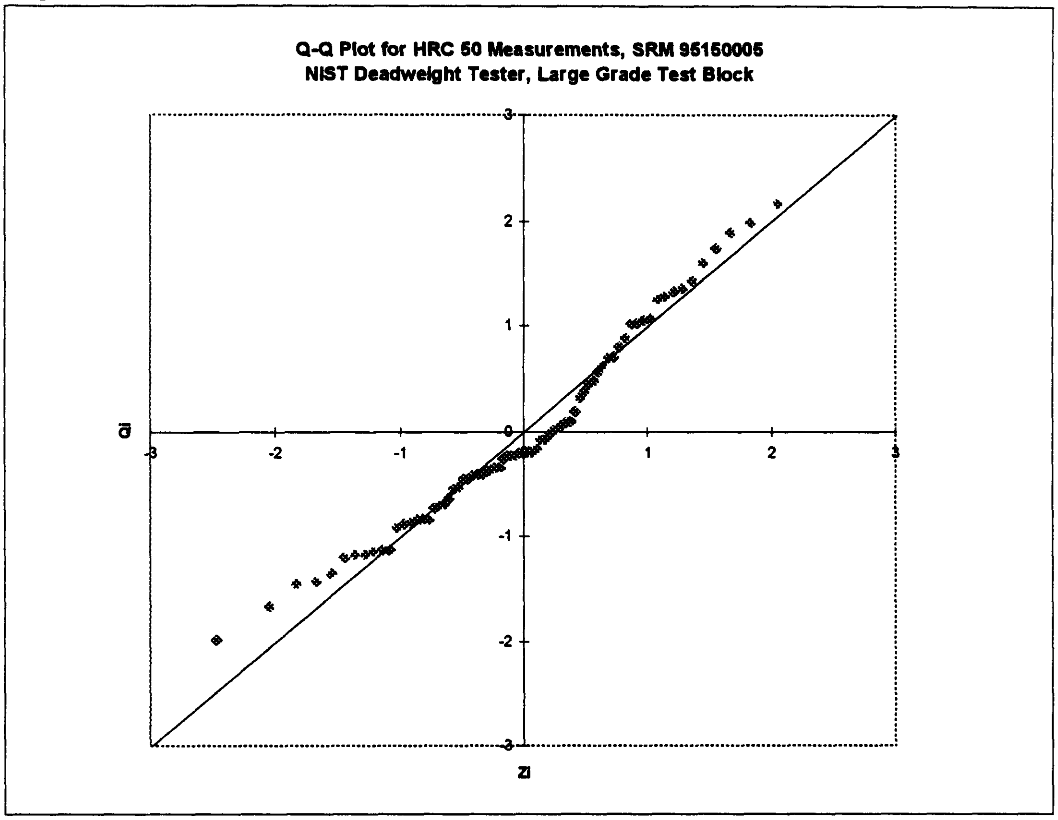


Figure F-8:

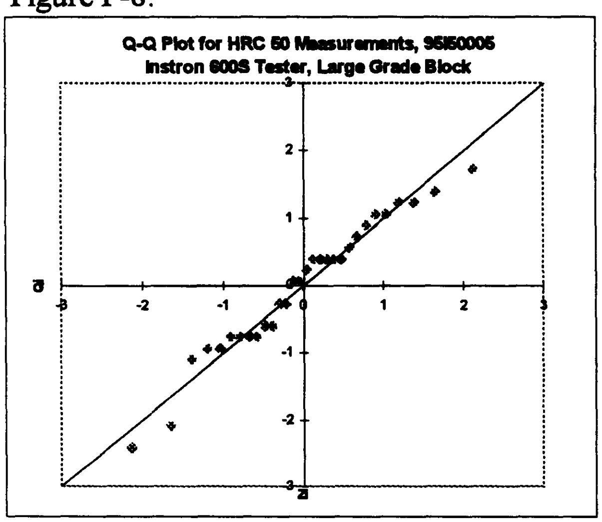
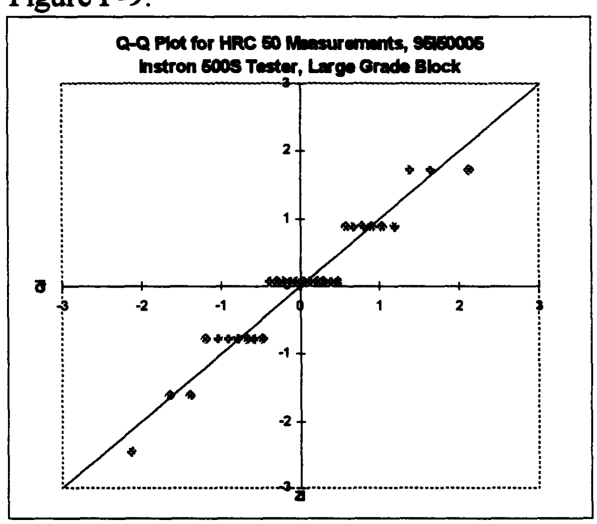


Figure F-9:



APPENDIX F Q-Q Plots for Normality Goodness-of-Fit

Figure F-10:

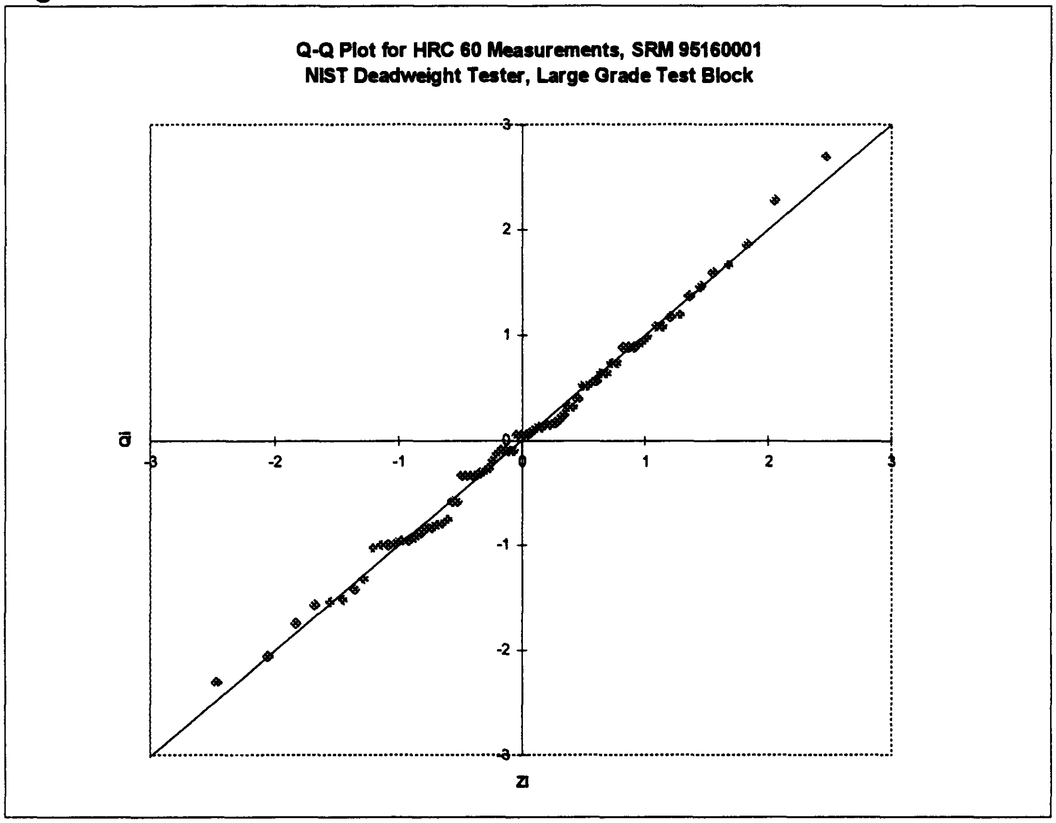


Figure F-11:

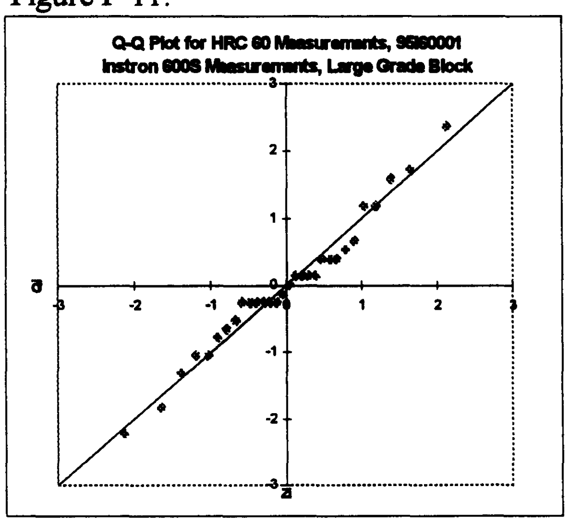
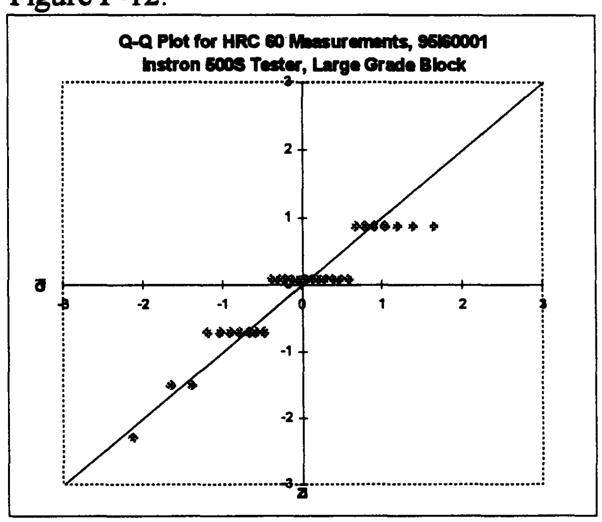


Figure F-12:



APPENDIX F

Figure F-13:

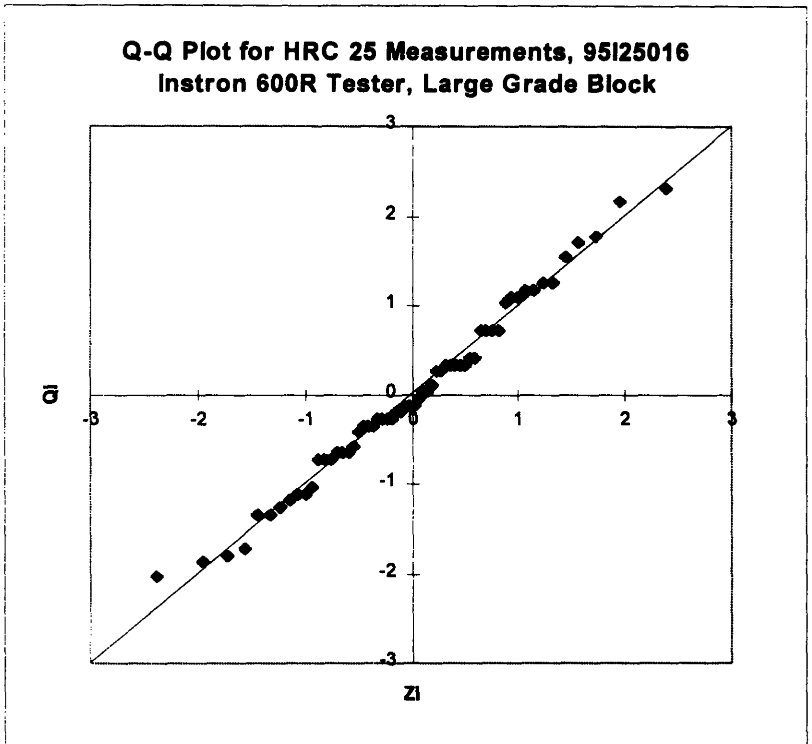
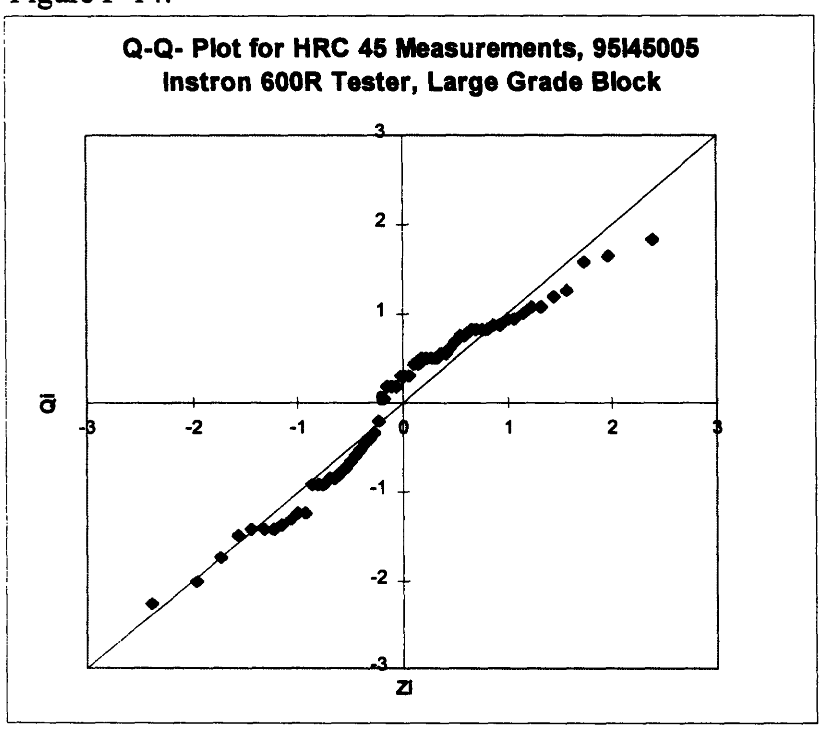


Figure F-14:



APPENDIX F

Figure F-15:

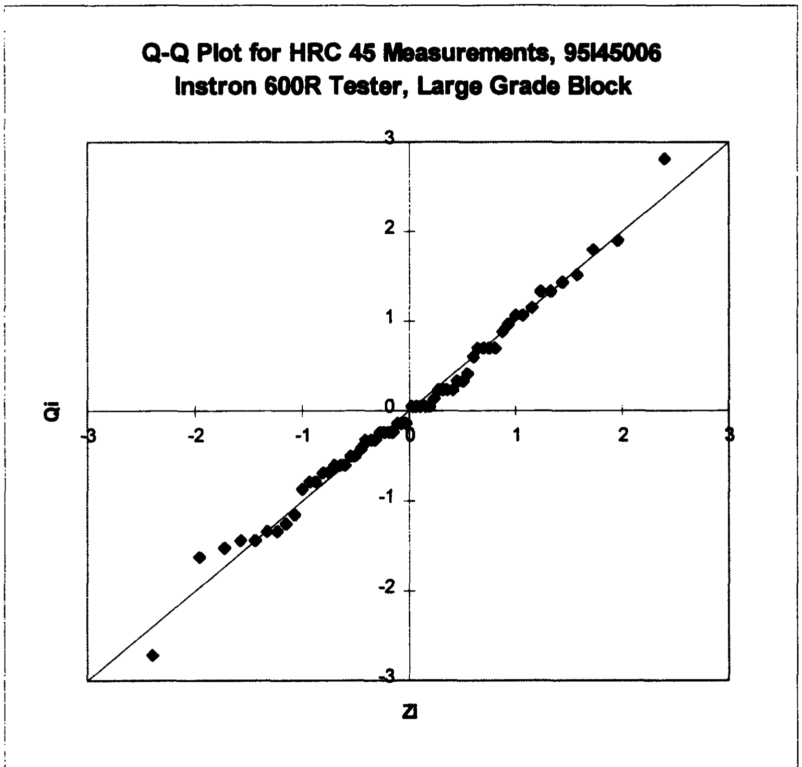
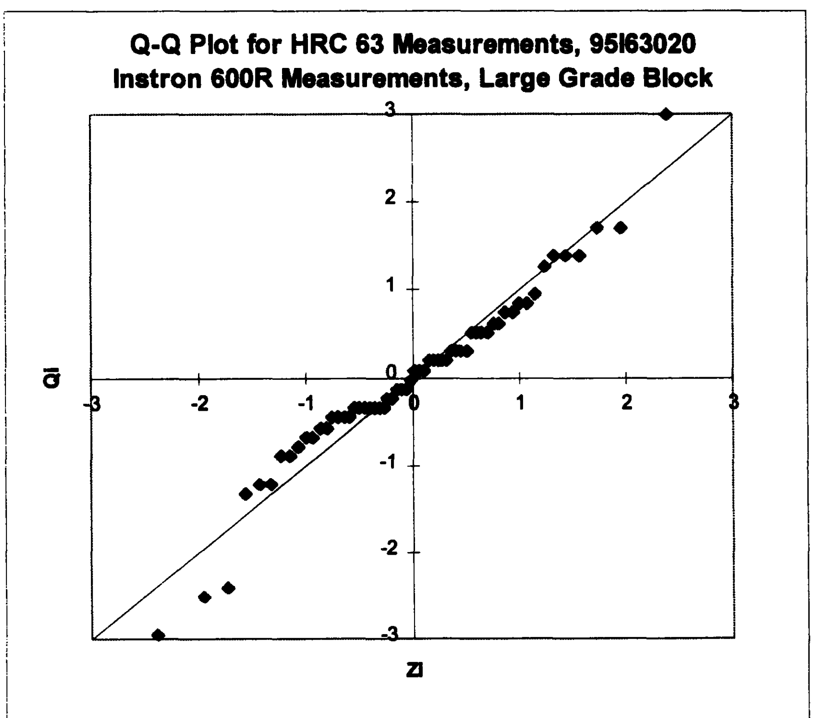


Figure F-16:



APPENDIX F **Q-Q Plots for Normality Goodness-of-Fit**

Figure F-17:

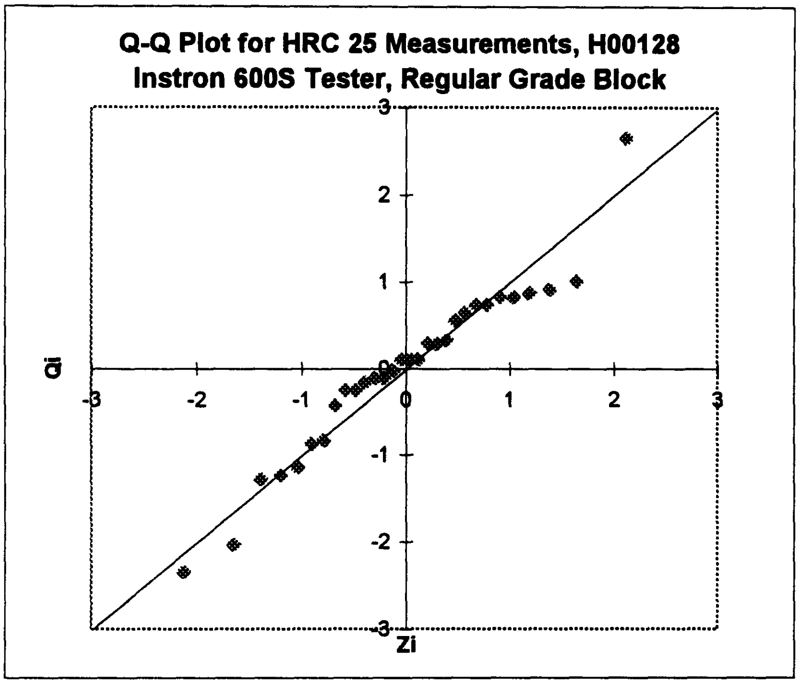
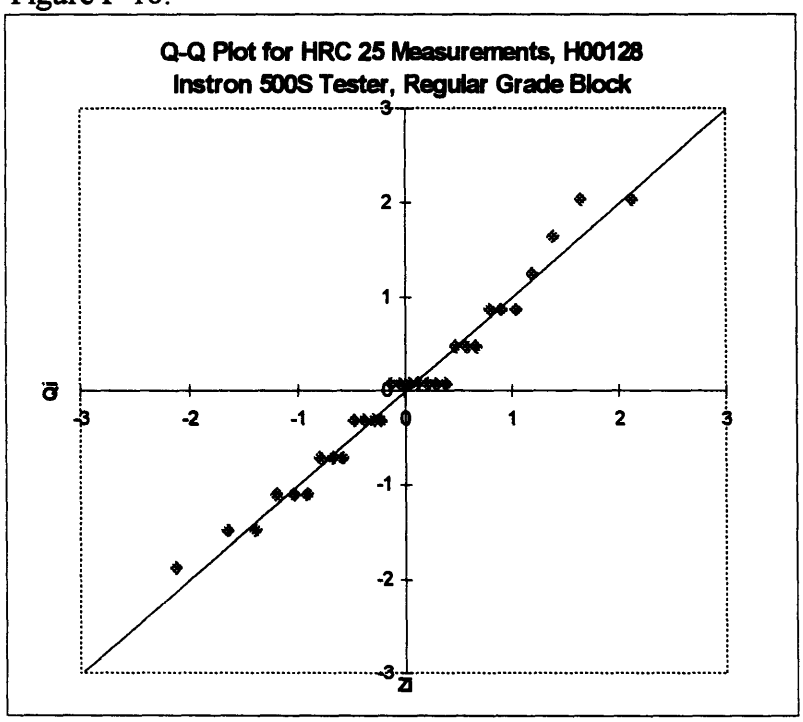


Figure F-18:



APPENDIX F **Q-Q Plots for Normality Goodness-of-Fit**

Figure F-19:

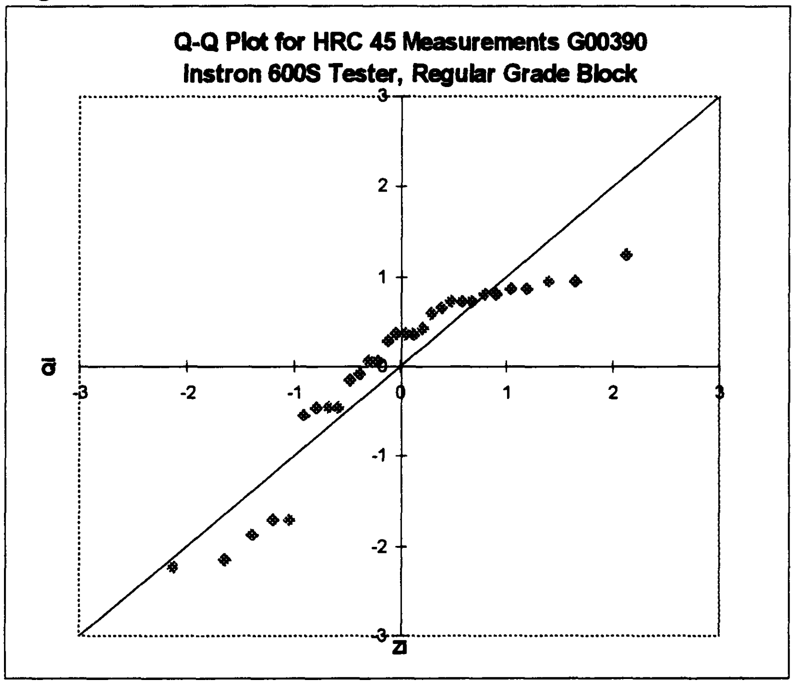
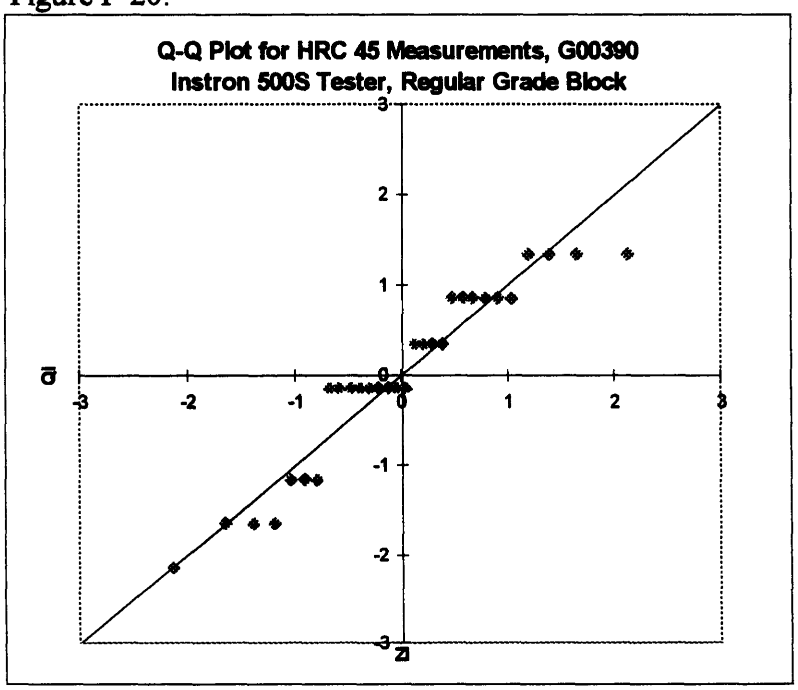


Figure F-20:



APPENDIX F **Q-Q Plots for Normality Goodness-of-Fit**

Figure F-21:

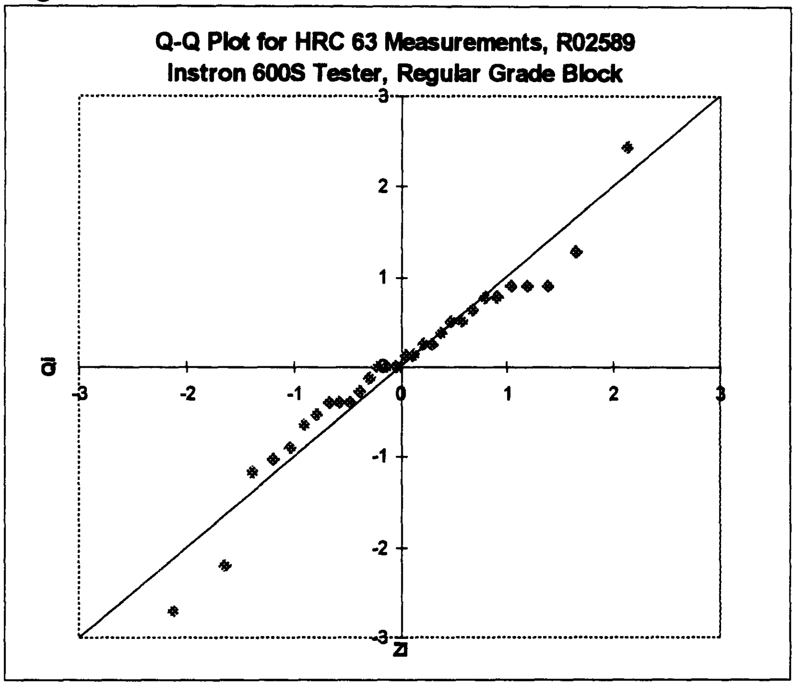
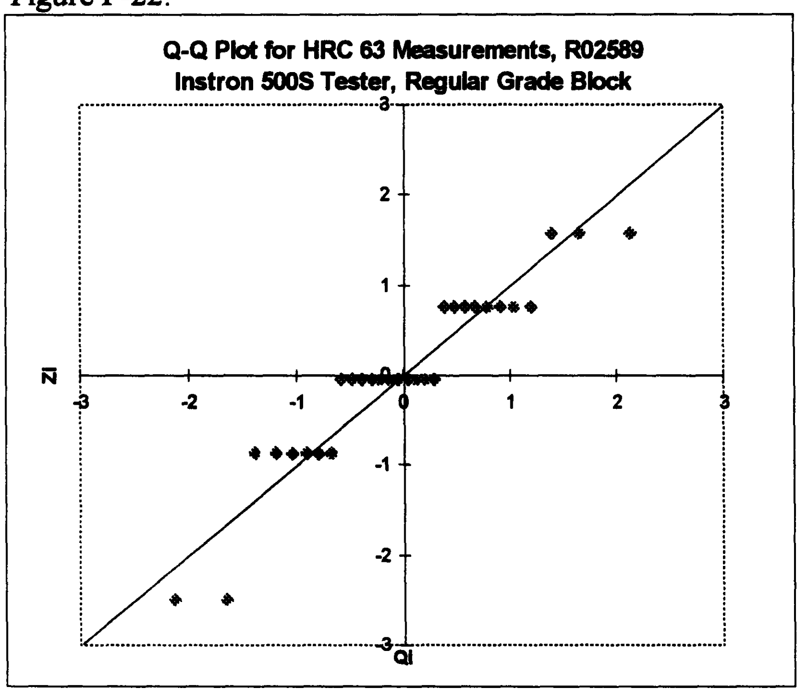


Figure F-22:



APPENDIX F Histograms for Normality Goodness-of-Fit

Figure F-23:

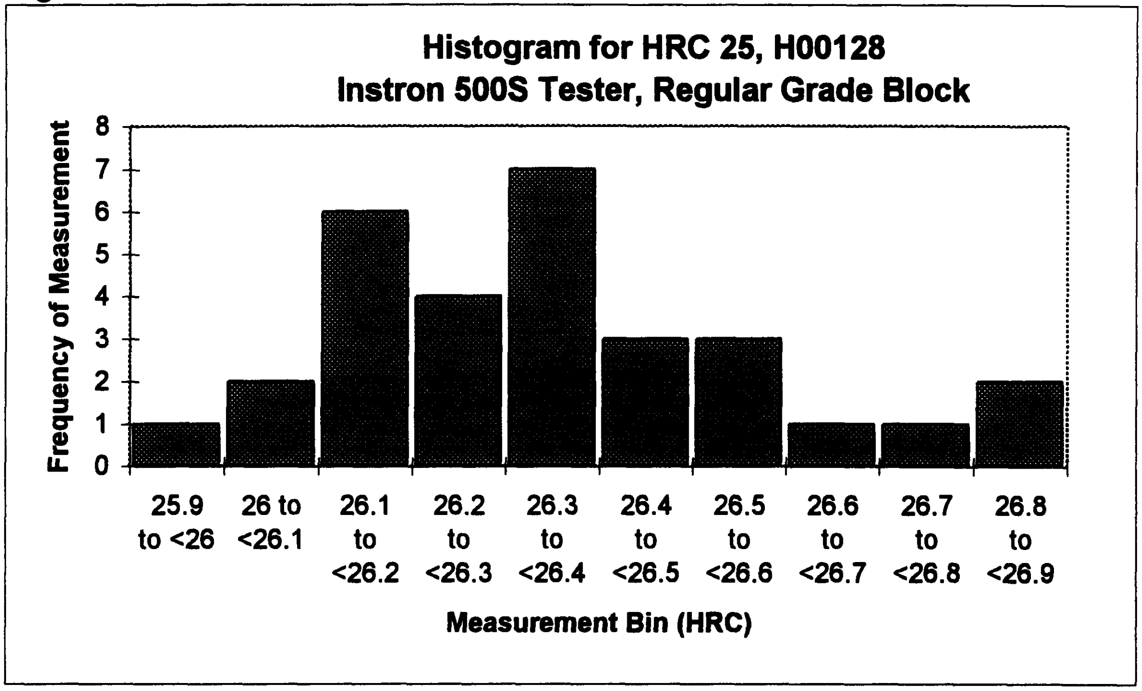
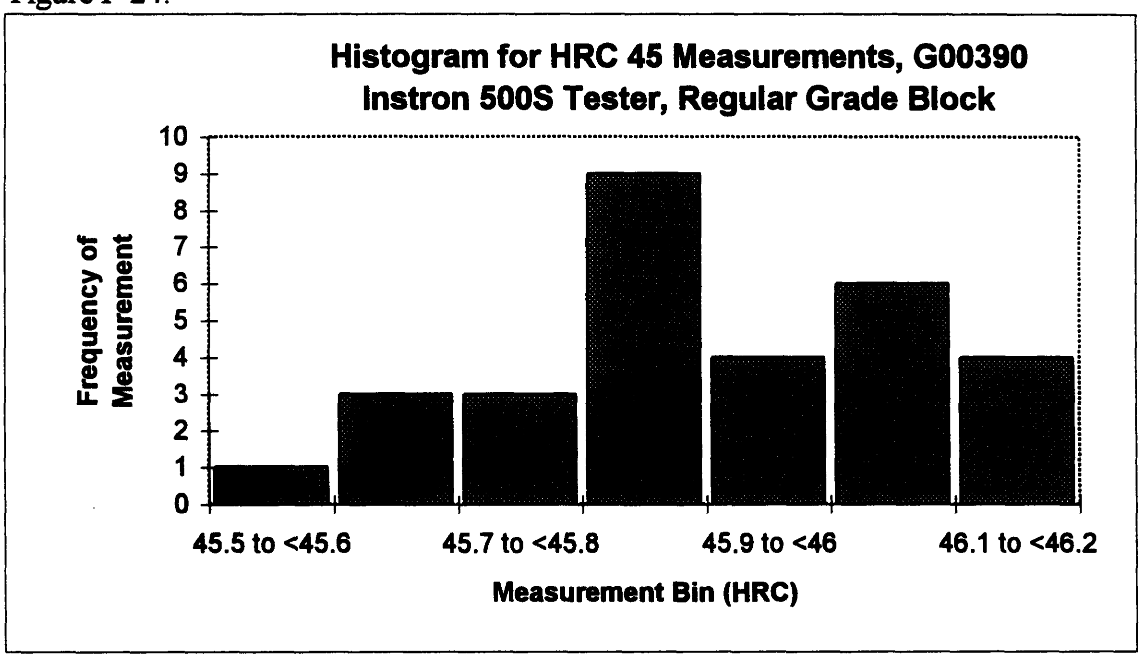
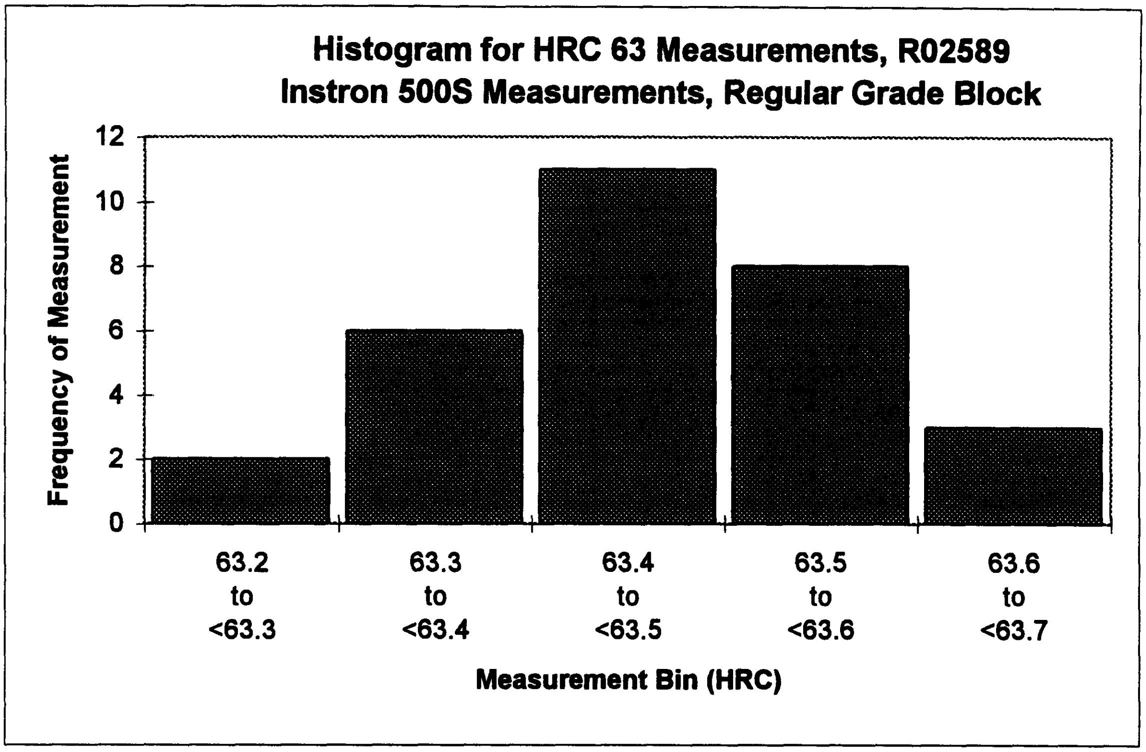


Figure F-24:



APPENDIX F **Histograms for Normality Goodness-of-Fit**

Figure F-25:



APPENDIX G Test block microstructures showing random distribution of steel iron-carbon phases and other constituents

Figure G-1: Typical test block microstructure, 500X, Large grade, 2% nital etch

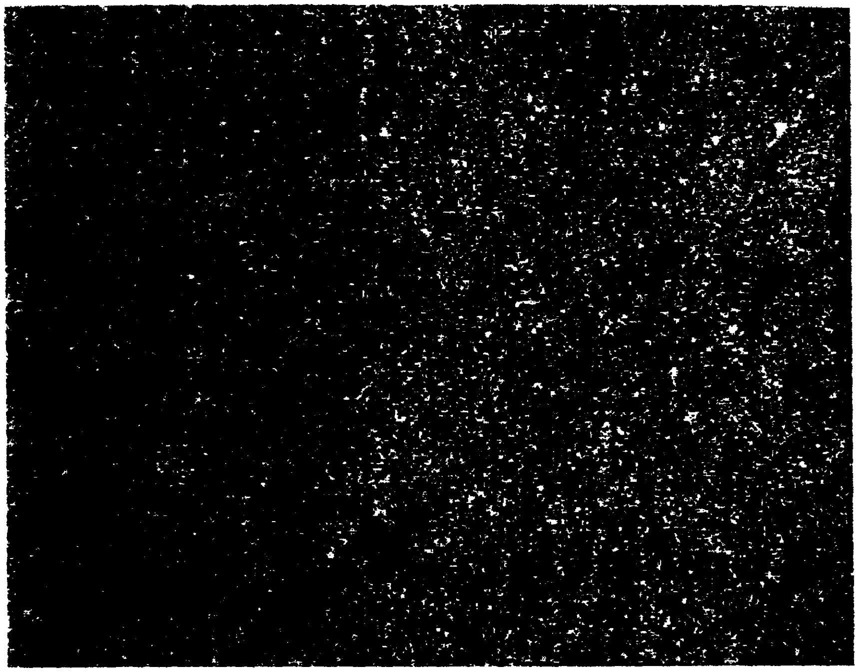
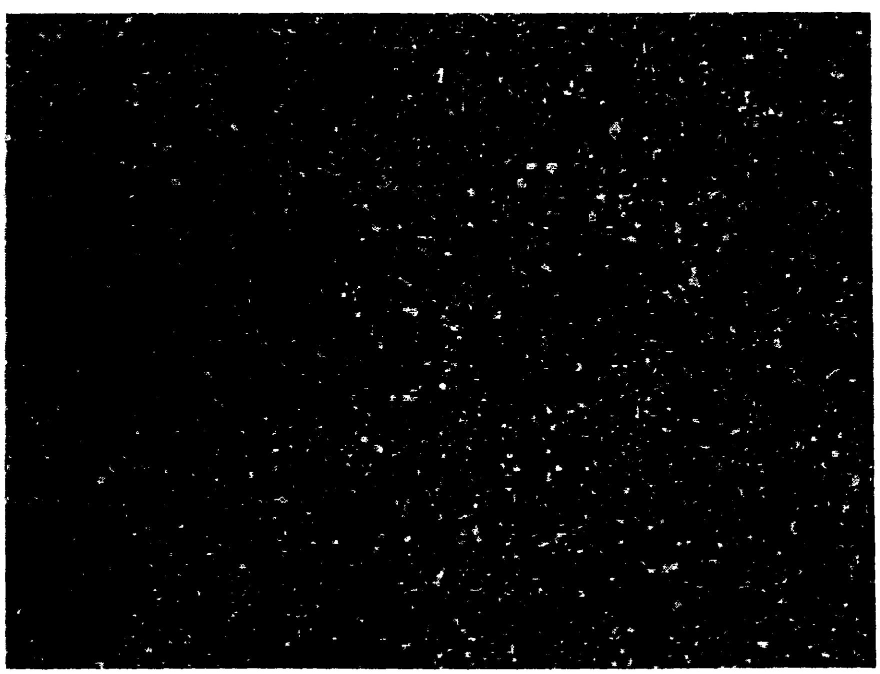
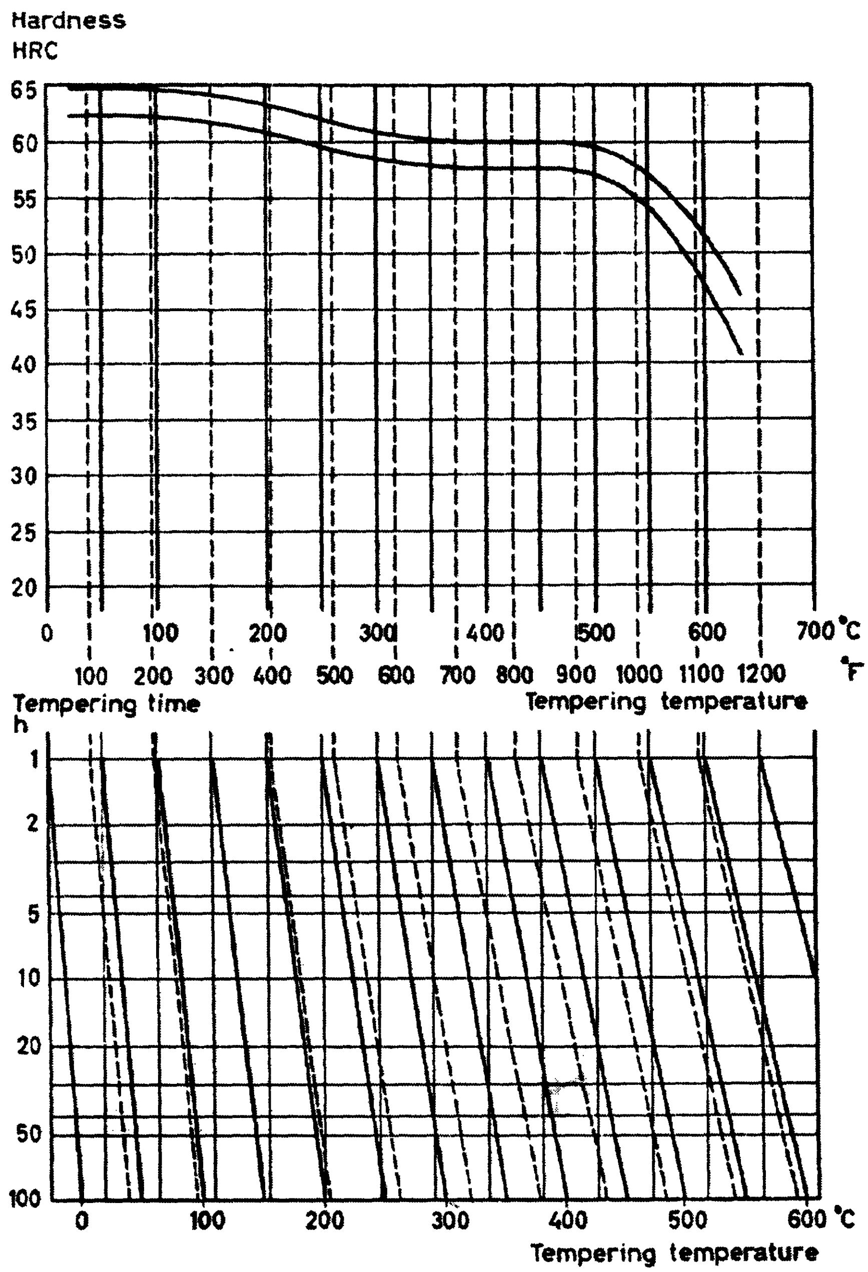


Figure G-2: Typical test block microstructure, 1250 X, Large grade, 2% nital



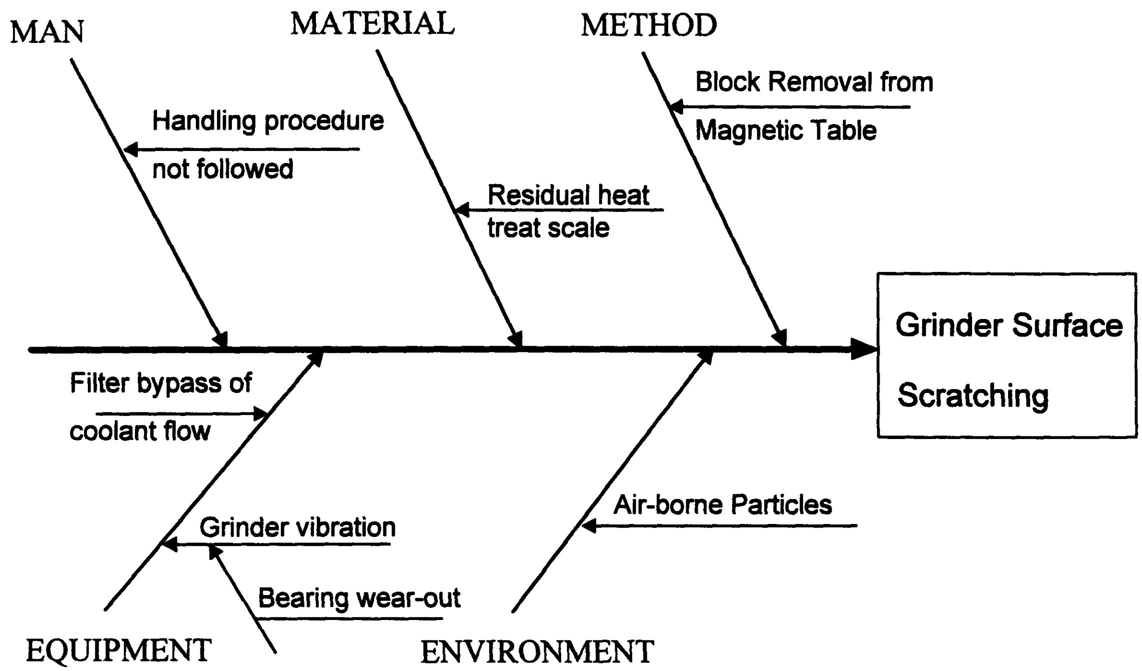
APPENDIX H Tempering Curve for Block Heat Treatment to Target Nominal Hardness Level

Figure H-1: Tempering curve showing interdependence between *time* and *temperature*



Ref.: Thelning, *Steel and its Heat Treatment* [20]

APPENDIX I Example Ishikawa Diagram for Root-Cause Investigation in the Test Block Manufacturing Process



Ref.: Shiba, *A New American TQM* [24]