Check for updates

NIST Special Publication 260
NIST.SP.260-269

Certification Report for SRM 2810a, 2811a, and 2812a:

Rockwell C Scale Hardness – Low, Medium, and High Ranges

Louise A. Ahure Powell
Carlos Beauchamp
Renato Reis Machado
Antonio Possolo

NIST NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

NIST Special Publication 260
NIST.SP.260-269

Certification Report for SRM 2810a, 2811a, and 2812a:

Rockwell C Scale Hardness – Low, Medium, and High Ranges

Louise A. Ahure Powell
Carlos Beauchamp
Renato Reis Machado

Materials Science and Engineering Division (MSED),
Material Measurement Laboratory (MML), National
Institute of Standards and Technology (NIST),
Gaithersburg, MD, USA

Antonio Possolo
Statistical Engineering Division (SED), Information
Technology Laboratory (ITL), National Institute of
Standards and Technology (NIST), Gaithersburg, MD,
USA

Renato Reis Machado
Mechanical Metrology Division (DIMEC), Scientific,
Industrial and Technology Metrology Directorate
(DIMCI), National Institute of Metrology, Quality and
Technology (INMETRO), Duque de Caxias - RJ, Brazil

This publication is available free of charge from:
https://doi.org/10.6028/ NIST.SP.260-269

August 2026

U.S. Department of Commerce
Howard Lutnick, Secretary

National Institute of Standards and Technology
Arvind Raman, NIST Director and Under Secretary of Commerce for Standards and Technology

NIST.SP.260-269
August 2026

Certain equipment, instruments, software, or materials, commercial or non-commercial, are identified in this paper in order to specify the experimental procedure adequately. Such identification does not imply recommendation or endorsement of any product or service by NIST, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

NIST Technical Series Policies

NIST Technical Series Publication Identifier Syntax

Publication History
Approved by the NIST Editorial Review Board on 2026-07-07

How to Cite this NIST Technical Series Publication
Ahure Powell LA, Beauchamp C, Machado RR, Possolo A. (2026) Certification Report for SRM 2810a, 2811a, and 2812a: Rockwell C Scale Hardness – Low, Medium, and High Ranges. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) NIST SP 260-269.
https://doi.org/10.6028/NIST.SP.260-269

Author ORCID iDs
Louise A. Ahure Powell: 0000-0002-4629-517X
Carlos Beauchamp: 0000-0002-4109-2824
Renato Reis-Machado: 0000-0002-8696-282X
Antonio Possolo: 0000-0002-8691-4190

Contact Information
louise.ahurepowell@nist.gov
carlos.deauchamp@nist.gov
rrmachado@inmetro.gov.br
Antonio.possolo@nist.gov

Abstract

The procedures used by the National Institute of Standards and Technology (NIST) to establish certified Rockwell C hardness (HRC) values for three Standard Reference Materials (SRMs): SRM 2810a (low range, nominal 25 HRC), SRM 2811a (medium range, nominal 45 HRC), and SRM 2812a (high range, nominal 63 HRC), are documented. These materials serve as the primary transfer standards of NIST's realization of the Rockwell C Hardness Scale. These standards are used for the calibration and verification of Rockwell C hardness testing machines, and they provide a direct link in the traceability chain through the International System of Units or SI-traceable force and depth measurements.

Certified values were obtained using the NIST Primary Hardness Standard Machine (PHSM) under conditions consistent with ASTM E18-25 (2025), ISO 6508-1 to -3 (2023), and the Consultative Committee for Mass and Related Quantities (CCM) Working Group on Hardness (WGH) definition of the Rockwell C hardness scale. The certification process included the direct verification of the test forces using calibrated force measuring instruments (load cells) and the calibration of the depth measuring system using laser interferometry with gauge block stacks, both traceable to the SI. It also incorporated a statistical evaluation of hardness measurements across multiple blocks spanning the low, medium, and high hardness ranges. Comparisons with historical NIST certifications and international key comparison results confirmed continuity of traceability.

The resulting certified values and expanded uncertainties provide industry and the hardness metrology community with reliable reference standards that ensure accuracy and global comparability of Rockwell C scale hardness measurements.

Keywords
Hardness testing; Rockwell C hardness (HRC); Standard Reference Material (SRM); Primary Hardness Standardizing Machine (PHSM); Traceability; Uncertainty.

Table of Contents

List of Tables

List of Figures

List of Source Codes

1. Introduction

Hardness testing plays a critical role in materials engineering, providing industry and research laboratories with a fast and reliable means of evaluating material hardness. To ensure global consistency, hardness scales must be realized (i.e., practically implemented in accordance with defined reference procedures) and maintained using highly accurate primary hardness standardizing machines operated at national metrology institutes (NMIs). The National Institute of Standards and Technology (NIST) Primary Hardness Standardizing Machine (PHSM) fulfills this role for the United States. As the foundation for the realization of the Rockwell, Vickers, Knoop, and Brinell hardness scales, the NIST PHSM supports the assignment of certified hardness values and underpins the traceability of hardness measurements in a well-defined metrological sense through documented reference procedures and controlled realization conditions.

Although Rockwell hardness values are expressed on a conventional scale and are not quantities of the International System of Units (SI), the certified values assigned to Rockwell C Standard Reference Materials (SRMs) are nevertheless metrologically traceable. According to the International Vocabulary of Metrology (VIM) [1], metrological traceability is defined as the property of a measurement result whereby it can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty (VIM 2.41). As noted in VIM 2.41, Note 1, this concept of traceability applies to conventional scales, including Rockwell C hardness, for which the reference is the defined realization of the scale (in the United States, as implemented by the NIST PHSM) rather than an SI unit. In this context, traceability does not imply direct linkage to the SI through a unit definition, but rather an unbroken chain of comparisons to the primary realization of the Rockwell C hardness scale, as described in NIST Technical Note 2156 [2].

This traceability chain is established through the calibrated realization of test forces, precise measurement of indentation depth, controlled timing, strict adherence to standardized test conditions specified in ASTM [3] and ISO [4, 5] standards, and the participation in Key Comparison exercises performed under the auspices of Consultative Committee of Mass and Related Quantities (CCM) Working group on Hardness (WGH), using their definition of the Rockwell C scale [6]. Certified SRMs produced using the NIST PHSM serve as primary transfer standards, enabling calibration laboratories to calibrate secondary-standard hardness blocks, and thereby extend the traceability chain to routine hardness testing in industrial, research, and quality-control environments. In doing so, the NIST hardness program enables both U.S. industry and international partners to maintain confidence in Rockwell C hardness measurements critical to manufacturing, quality assurance, and materials research.

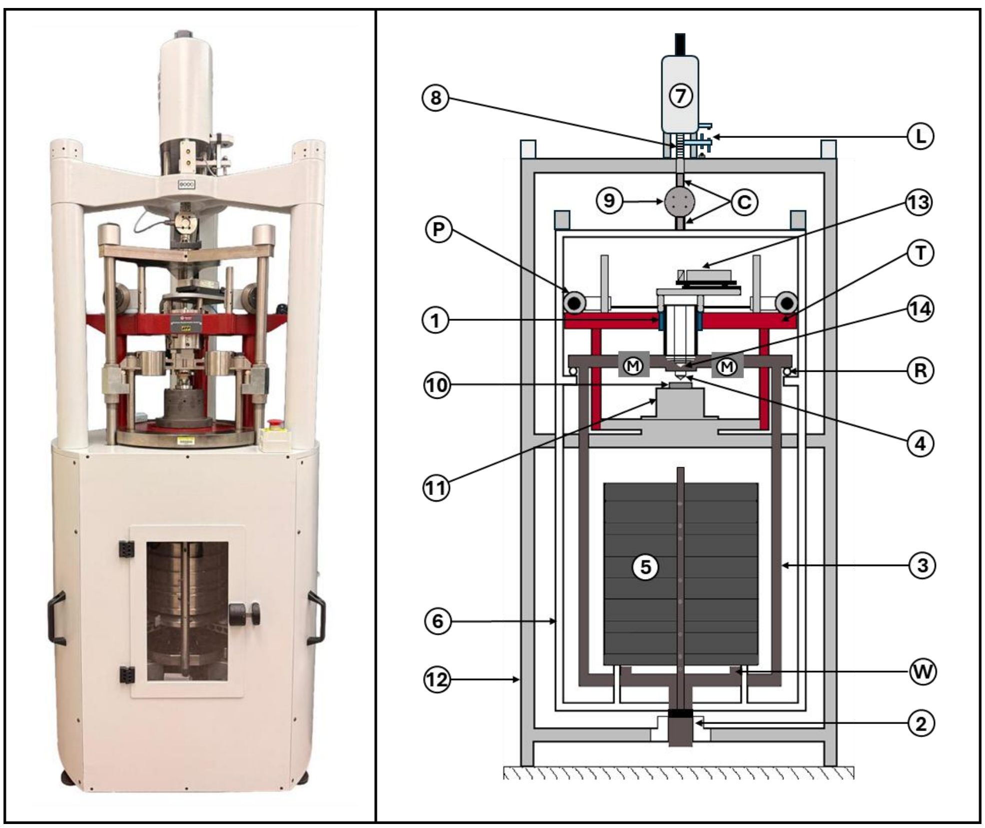
2. The NIST Primary Hardness Standardizing Machine (PHSM)

2.1. History of the NIST PHSM

Over the past several decades, NMIs have developed and refined primary hardness standardizing machines to ensure reliable and reproducible realization of hardness scales [7-17]. A related machine was custom-built for NIST (USA) by LTF SpA based on an updated version of the Istituto di Metrologia "G. Colonnetti" (IMGC, now the Istituto Nazionale di Ricerca Metrologica, INRiM) design [7], which demonstrated measurement repeatability better than ± 0.1 HRC [18].

Conceived and constructed in the late 1980s, the original NIST PHSM underwent successive upgrades in the early 1990s, introducing modern electronics, advanced electromechanical systems, and software-based control resulting in improved measurement control and repeatability. In 2017, NIST acquired its current PHSM with enhanced measurement capabilities, further strengthening the system's overall performance.

The current NIST PHSM is designed to realize multiple hardness scales with the highest metrological accuracy and provides full capability across the Rockwell scales, defined in ASTM [3] and ISO [4, 5] standards, as well as the Vickers and Brinell scales (see **Figure 1a**).



(a) (b)

Figure 1. NIST Primary Hardness Standardizing Machine (PHSM) for Rockwell, Vickers, and Brinell scales: (a) mechanical assembly associated with force application, indentation formation, and indenter depth measurement (b) schematic diagram showing the principal mechanical components.

2.2. Force Application and Depth Measuring System

Figure 1b shows the principal mechanical components of the NIST PHSM involved in force application and indentation depth measurement. Two air bearings (1, 2) are aligned with the axis of the indenter frame (3). The upper air bearing (1) enables frictionless vertical motion, while the lower air bearing (2) provides resistance to tilting moments and maintains alignment.

The indenter frame (3) consists of an upper cross plate supporting the indenter (4) and three tie-rods connected to a lower plate that supports the stainless-steel masses (5) used to generate the major test force. The frame weighs 29.42 N (3 kgf). For the Rockwell C scale, where the preliminary force must be 98.07 N (10 kgf), three additional masses (M), equally spaced on the upper plate, provide the remaining force required to achieve the preliminary force, each contributing approximately 22.85 N (2.33 kgf).

A counterweight three-pulley system (P) (only two pulleys are visible in **Figure 1**) offsets part of the frame weight during other test cycles (e.g., a Vickers microhardness test with a 1 kgf force, 9.807 N, also called HV1). The pulley system is disabled during Rockwell C testing, allowing the full frame weight and additional masses (M) to provide the preliminary force.

The lower plate of the indenter frame (3) forms a basket supporting the mass stack (5) at total force. Initially, the masses are supported by an outer moving frame (6) driven by a servomotor (7) located in the upper portion of the machine. Motion is transmitted through a high precision ball screw mechanism (8), and safety limit switches (L) interrupt motor power if the frame reaches either travel limit.

A force measuring instrument (9), with a capacity of 2941.99 N (300 kgf), is mounted between two elastic couplings (C) between the screw and the outer frame (6). Three equally spaced adjustable rests located on the outer frame (6) support spacer balls (R), which in turn support the indenter frame (3) when the indenter is not in contact with the block being measured.

The hardness reference block (10) is supported by an anvil (11) mounted on a rigid stationary structure (i.e., the base of the instrument) that carries the indentation measurement system and is attached to the main stationary frame (12).

Indentation depth is measured with a laser interferometer system (13) mounted on the cross plate (T) of the tester stationary frame (12). The system includes a laser head, polarized separator, and a fixed corner cube. A moving corner-cube reflector (14), attached to the indenter frame (3) above the indenter (4), measures the vertical displacement of the basket and thus the indentation depth necessary for hardness calculation.

At the start of an indentation cycle, the interferometer is reset, and the servomotor (7) lowers the outer frame (6) at a controlled velocity of approximately 0.2 mm/s. When the indenter contacts the hardness reference block (10), the indenter frame (3) separates from the adjustable rests (R), transferring the preliminary force to the indenter.

As the outer frame continues to descend, the mass stack (5) is transferred onto three loading pins (W) fastened to the indenter frame (3), applying the major test force. Any lateral forces generated during force transfer are absorbed by the air bearings (1, 2). Because the centers of gravity of the masses are closely aligned with the indenter axis, operation is possible even with the air supply to the bearings is turned off.

After the prescribed dwell time at the total test force, the servomotor reverses direction, and the selected mass stack (5) is smoothly removed from the indenter frame (3). The indentation depth is then obtained from the interferometer measurement, and the indenter is lifted from the test surface, completing the test cycle.

2.3. Advancements in the NIST Primary Hardness Standardizing Machine

Building upon its long history of development, the current NIST PHSM represents the thirteenth advancement in this lineage of precision instruments, introducing major improvements in design, measurement capability, and test cycle control that establish the state of the art in hardness metrology. It integrates fully customized software for enhanced control across all testing scales, alongside expanded deadweight and cycle-control capabilities to accommodate Rockwell, Vickers, and Brinell hardness scales. This system is currently used in the production of new SRM batches, resulting in a revised designation for the Rockwell C SRMs from 2810, 2811, and 2812 (low, medium, and high hardness ranges) to 2810a, 2811a, and 2812a, respectively. The suffix “a” denotes a new realization of the scale based on the updated NIST PHSM and an associated change in the certification methodology, including a revised statistical analysis (e.g., presenting different modes of certification and using 13 indentations instead of 7 for results analysis).

A distinguishing feature of the current NIST PHSM is its integration of a new compact, high-precision laser interferometric system for indentation depth measurement. Combined with its existing deadweight system for force application, the NIST PHSM ensures exceptional reproducibility in hardness scale realization. The machine’s high stiffness and isostatic structural design minimize deformation and error sources, while updated, flexible software provides precise control over key test cycle parameters, such as dwell times and rates of force application and removal.

With these capabilities, in the United States, the NIST PHSM achieves unmatched accuracy and repeatability in the realization of hardness scales. It serves not only as the foundation for NIST’s production of hardness SRMs, but also as a benchmark system that strengthens international standardization in hardness metrology.

The robustness of the measurements was established through a comprehensive series of verifications, described in sections **3.1.2** and **5**, and performed on the NIST PHSM prior to the SRM certifications.

3. Experimental Methods

3.1. Materials

3.1.1. Standard Reference Materials (SRMs)

The hardness levels represented in the three SRMs being certified (SRM 2810a, SRM 2811a, and SRM 2812a) were selected to cover the practical Rockwell C hardness range of interest and meet the requirements of ASTM [3] and ISO [4, 5] standards:

- Low hardness: 25 HRC nominal (SRM 2810a)
- Medium hardness: 45 HRC nominal (SRM 2811a)
- High hardness: 63 HRC nominal (SRM 2812a)

These levels ensure that the certified SRMs provide traceability across a broad range of hardness values relevant not only to industrial and research applications, but also to calibration laboratories that certify working-standard hardness blocks used for routine machine verification. In this way, the selected nominal levels support continuity of the Rockwell C traceability chain – from the NIST PHSM to secondary laboratories, and ultimately to end-users in testing facilities, manufacturing plants, and quality control environments.

Blocks at each nominal level were subjected to the full certification process, including in-process control checks, seating, and randomized indentation patterns, as described in this document.

The SRM hardness test blocks are circular, with the nominal dimensions shown in **Table 1**. The NIST logo and an engraved circle – to delineate the boundary of the designated measurement area – are present on the testing surface of the block. A schematic representation of SRM hardness test block is shown on **Figure 2**.

Table 1. Nominal dimensions of SRM hardness test block.

Feature	Nominal Dimensions (mm)
Block diameter	64
Block thickness	15
Engraved circle diameter	57

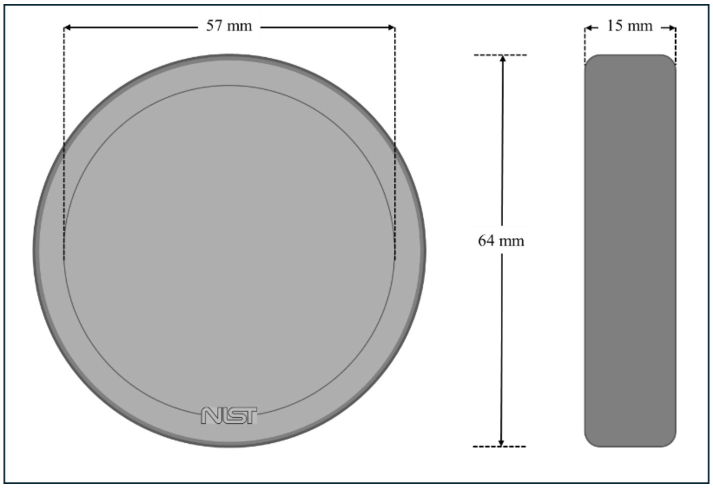


Figure 2. Schematic of a NIST SRM hardness block.

3.1.2. Indirect Verification: Control Blocks for In-Process Monitoring

In addition to the SRMs, dedicated control blocks were used during the certification process to provide ongoing verification of system performance. These control blocks were selected from the same batch as the SRMs and span the same range of hardness values, representing low, medium, and high Rockwell C hardness levels. Specifically, control blocks with nominal hardness values of 25 HRC, 45 HRC, and 63 HRC were employed, corresponding to the ranges of SRM 2810a, SRM 2811a, and SRM 2812a, respectively.

The control blocks are circular and manufactured to the same nominal dimensions as the SRM blocks, ensuring comparable testing conditions. Each block was subjected to repeated hardness measurements throughout the course of the certification process. By monitoring the results obtained on these control blocks, the reproducibility and stability of the NIST PHSM were continuously assessed, independent of the certification measurements performed on the SRM candidates.

The use of low-, medium-, and high-level control blocks provided confidence that the NIST PHSM maintained consistent performance across the entire Rockwell C range of interest. These results form the basis of the in-process control checks described in the following section.

3.1.2.1. Day-to-Day Reproducibility

All measurements were performed under controlled environmental conditions, with temperature and relative humidity continuously monitored using calibrated sensors that satisfy the applicable standards [3, 5] and the CCM-WGH's definition of the Rockwell C scale [6]. Throughout the SRM certification process, the laboratory temperature was maintained at 23 °C ± 1 °C.

To ensure continued stability and reliability throughout the SRM certification process, in-process control checks were conducted in parallel with the certification measurements. These checks involved repeated hardness measurements performed on the control blocks described in **3.1.2.**, representing nominal Rockwell C hardness levels of 25 HRC, 45 HRC, and 63 HRC.

The purpose of these in-process control checks was to monitor the day-to-day reproducibility of the NIST PHSM, and to verify that its performance remained within the expected repeatability limits defined for primary hardness standardizing machines over the duration of the certification.

Each day prior to SRM testing, the three control blocks were indented under the same test parameters (defined in **4. Testing Methods and Parameters**) used for the certification of the SRM candidate blocks. For each block, the daily mean Rockwell C hardness and corresponding standard deviation were calculated, and these daily mean values were then tracked across all measurement days to assess day-to-day reproducibility. These in-process data provided a direct means of detecting any deviation or drift in machine performance that could affect the certification results. If variations outside the allowable control limits [3, §A2.6.3 Table A2.6, and 5, §D.2] had been observed, corrective actions, such as reseating the blocks, verifying the depth measuring system or applied force calibrations, or confirming environmental stability, would have been implemented before continuing certification measurements.

Across all three hardness levels, the NIST PHSM exhibited highly consistent and stable performance. The standard deviation of the daily mean hardness values was:

- ± 0.095 HRC for the 25 HRC control block,
- ± 0.057 HRC for the 45 HRC control block, and
- ± 0.038 HRC for the 63 HRC control block.

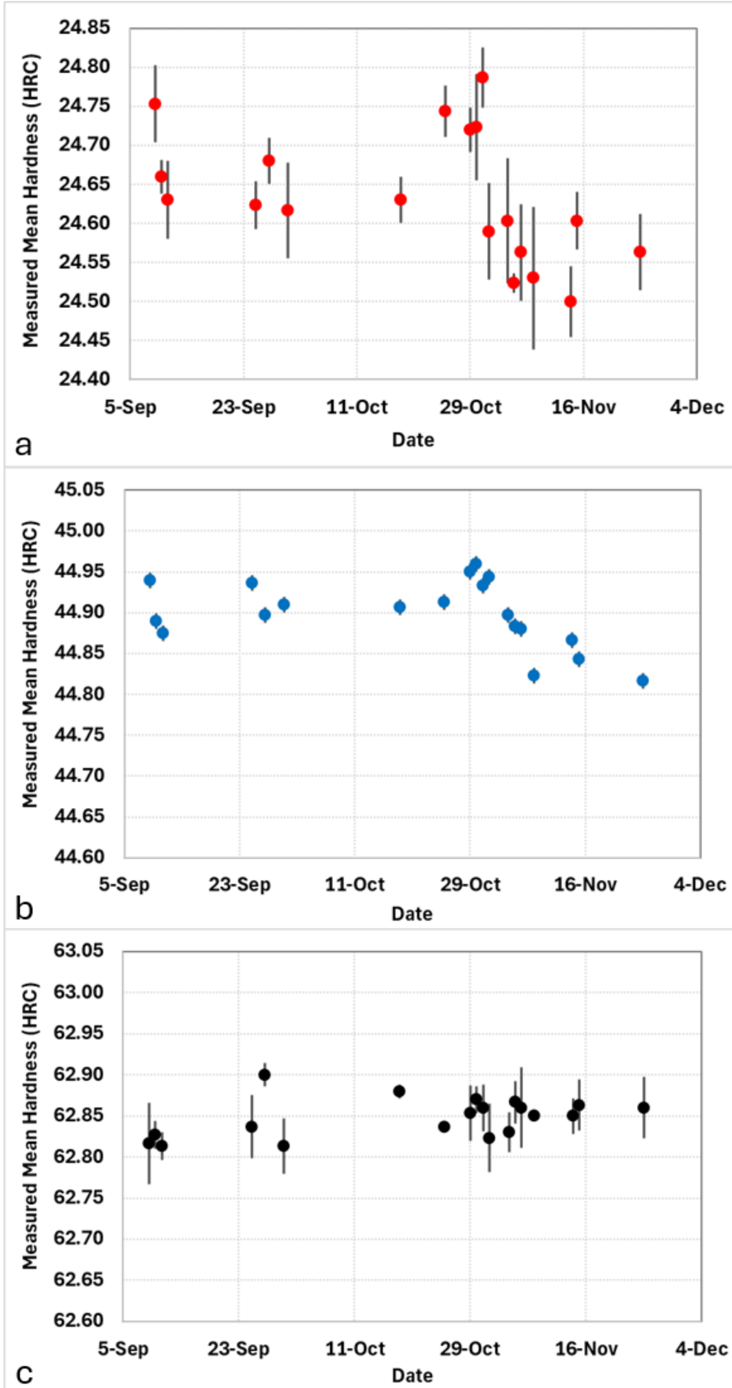


Figure 3. Day-to-day reproducibility of three Rockwell C scale control blocks used during SRM certification. Each point represents the daily mean hardness value, with error bars indicating the corresponding standard deviation. The standard deviation of the daily mean hardness values was: (a) 0.095 HRC for the 25 HRC control block, (b) 0.057 HRC for the 45 HRC control block, and (c) 0.038 HRC for the 63 HRC control block. All three plots use identical vertical axis scaling (same hardness scale and unit length).

These values reflect the day-to-day variability of the daily means at each hardness level. For each day, the plotted error bars (**Figures 3a to 3c**) correspond to the one-sigma within-day standard deviation of the repeated measurements obtained on that control block.

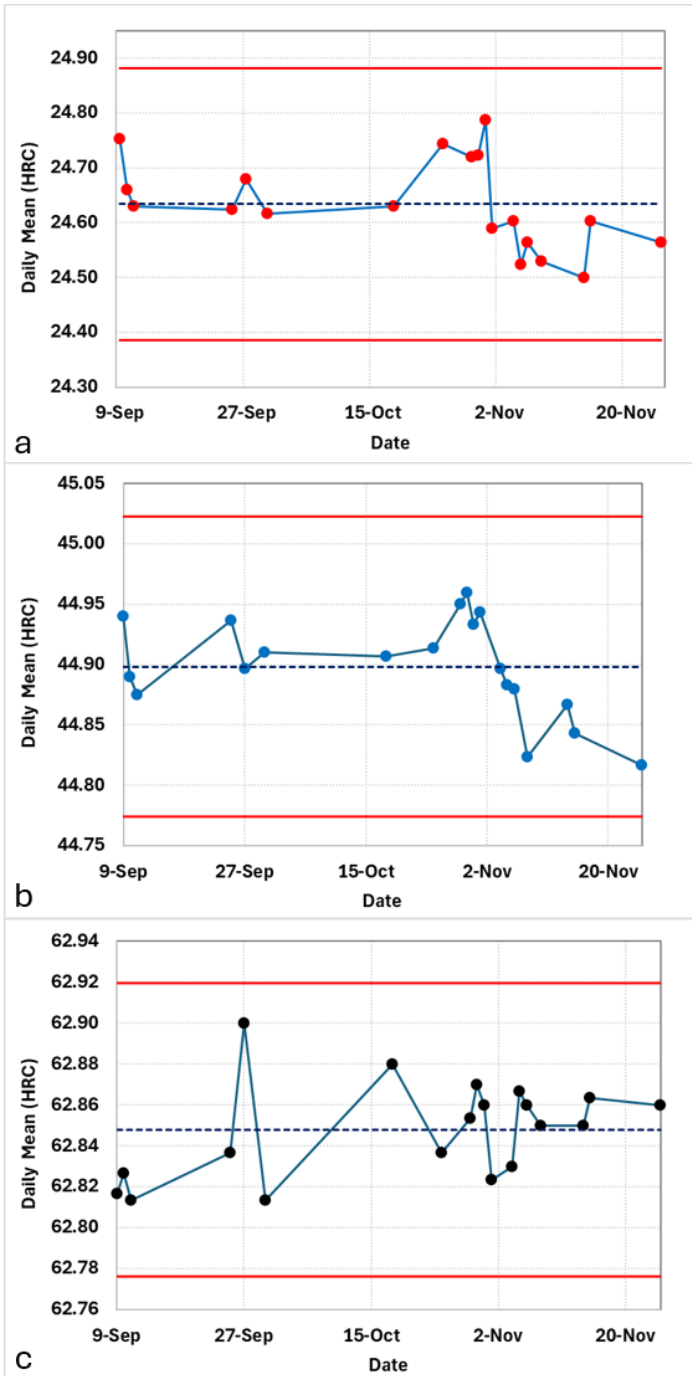


Figure 4. Day-to-day reproducibility control charts for the NIST PHSM, based on daily mean Rockwell C hardness values measured on three control blocks, at nominal hardness levels of (a) 25 HRC, (b) 45 HRC, and (c) 63 HRC. The center line represents the overall mean hardness value for each control block, and the upper and lower control limits correspond to ± 3 standard deviations of the observed daily means.

At the low hardness level (25 HRC), daily means of the control block remained tightly clustered with a day-to-day standard deviation of approximately ± 0.095 HRC, indicating stable force application and depth measurement performance in the low-hardness regime. At the medium hardness level (45 HRC), the day-to-day variation decreased to ± 0.057 HRC, consistent with the reduced influence of surface and environmental effects at mid-range hardness. At the high hardness level (63 HRC), the control block repeatability was excellent, with a low standard deviation of ± 0.038 HRC for the daily means, demonstrating robust force control and indenter stability at higher loads.

No outliers or systematic drifts were detected, and no recalibration or additional environmental corrections were required during the certification period.

Day-to-day reproducibility of the NIST PHSM was also evaluated using control charts based on daily mean Rockwell C hardness values obtained from the three control blocks spanning the nominal hardness range of the SRM candidates. **Figures 4a to 4c** show that all data remained within the allowable control limits [3, §A2.6.3 Table A2.6, and 5, §D.2], indicating stable NIST PHSM performance throughout the SRM certification.

3.1.2.2. Day-to-Day Reproducibility

a. System Stability

The observed day-to-day standard deviations of the daily means (± 0.095 HRC, ± 0.057 HRC, and ± 0.038 HRC for the 25 HRC, 45 HRC, and 63 HRC control blocks, respectively) are consistent with NIST PHSM characterization data [19], and confirm that the force application, depth measurement, and environmental controls remained stable throughout the certification process.

b. Load-Dependent Sensitivity

A decrease in variability with increasing hardness was observed, which is consistent with the expected load-sensitivity behavior of Rockwell C testing.

c. Metrological Confidence

Consistency across low, medium, and high hardness regimes reinforces the traceability and reproducibility of the NIST PHSM, validating its suitability for Rockwell C SRM certification without the need for intermediate recalibration. The in-process control strategy, based on daily means and associated within-day standard deviations for dedicated control blocks, provides a robust framework for real-time performance monitoring in current and future certification cycles.

4. Testing Methods and Parameters

4.1. Rockwell C Hardness Test Cycle

The NIST Rockwell C hardness test cycle is governed by harmonized requirements defined across standards [3-5] and especially the definition of the Consultative Committee for Mass and Related Quantities Working Group on Hardness (CCM-WGH) of the International Committee of Weights and Measures (CIPM) [6]. These standards collectively establish the consensus for force application sequence, dwell times, approach velocities, and measurement procedures required to achieve reproducible and internationally comparable results.

The cycle begins with the application of a preliminary force (F_0). An initial depth reading (h_i) is taken after the preliminary dwell period ($\frac{t_{pa}}{2}+t_{pd}$). The force is then incrementally increased to the total test force (F_T) over a defined loading interval (t_{aa}), typically between 1 s to 8 s. To minimize dynamic effects and to comply with the requirements of the mean indentation velocity during the final stage of the additional force application (between c_1F_T to c_2F_T) as specified on the documentary standards and the HRC scale definition, the rate at which the force is increased is adjusted once during the period t_{aa}. The constants $c_1$and c_2 are defined in **Tables 2** and **3**, and the associated approach mean velocity ($\overline{v_{fa}}$) is specified in the respective standards. ASTM requires 20 µm/s ≤ $\overline{v_{fa}}$ ≤ 40 µm/s [3, §A2.3.1.6], ISO permits mean values in the range of 15 µm/s ≤ $\overline{v_{fa}}$ ≤ 40 µm/s [5], and CCM-WGH establishes a reference value of 30 µm/s for the mean velocity between c_1F_T and c_2F_T [6]. Once F_T is reached, a total force dwell time (t_{td}) is required. The force is then reduced back to F_0 within a maximum of 2 s (t_{ar}). A recovery dwell (t_{rd}) follows to account for elastic recovery, after which the final depth (h_f) is recorded relative to the initial reference depth (h_i). The Rockwell C hardness number (HRC) is then calculated from the depth difference, shown in **Equations (1)** and **(2)**:

$$h = |h_f - h_i| \quad \textbf{(1)}$$

$$HRC = 100 - \frac{h}{0.002} \quad \textbf{(2)}$$

This unified framework, established in [3, 5-6], ensures that NIST SRM-certified values are consistent with the traceable international definition of the Rockwell C hardness scale.

Table 2. NIST Rockwell C parameters used for SRM certification.

Parameters (NIST SRM Requirements)	**Target Nominal Values**
Preliminary force ($\boldsymbol{F_0}$)	98.07 N
Total Force ($\boldsymbol{F_T}$)	1471 N
Constants ($\boldsymbol{c_1}$ & $\boldsymbol{c_2}$) for determining the mean indenter velocity during the final additional test force application (% $\boldsymbol{F_T}$)	Compliance with: ASTM: $c_1 = 0.6$, $c_2 = 1$ ISO/CCM-WGH: $c_1 = 0.8$, $c_2 = 0.99$
Preliminary force application dwell time ($\boldsymbol{t_{pa}}$)	< 1 s
Time duration ($\boldsymbol{t_{pd}}$) before initial depth reading ($\boldsymbol{h_i}$)	2.5 s
Additional force application time, from $\boldsymbol{F_0}$ to $\boldsymbol{F_T}(\boldsymbol{t_{aa}})$	5.4 s to 5.8 s
Approach mean indenter velocity ($\overline{\boldsymbol{v_{fa}}}$) between $\boldsymbol{c_1F_T}$ and $\boldsymbol{c_2F_T}$	30 µm/s
Total force dwell time ($\boldsymbol{t_{td}}$)	5 s
Time ($\boldsymbol{t_{ar}}$) to return to $\boldsymbol{F_0}$	1.8 s
Dwell time ($\boldsymbol{t_{rd}}$) for elastic recovery	4 s
Temperature	23 °C ± 1 °C

Table 3. Comparison of Rockwell C test cycle requirements across ASTM [3] and ISO [5] standards and the CCM-WGH's definition [6].

<table>
<tr><th>Parameters</th><th>ASTM E18-25</th><th>ISO 6508-3:2023</th><th>CCM-WGH (2021)</th></tr>
<tr><td>Preliminary Force ($\boldsymbol{F_0}$)</td><td colspan="2">98.07 N</td><td>98.0665 N</td></tr>
<tr><td>Total Force ($\boldsymbol{F_T}$)</td><td colspan="2">1471 N</td><td>1470.9975 N</td></tr>
<tr><td>Constants ($\boldsymbol{c_1}$ & $\boldsymbol{c_2}$) for determining the mean indenter velocity during the final additional test force application (% $\boldsymbol{F_T}$)</td><td>$c_1 = 0.6$,
$c_2 = 1$</td><td colspan="2">$c_1 = 0.8$,
$c_2 = 0.99$</td></tr>
<tr><td>Preliminary force application dwell time ($\boldsymbol{t_{pa}}$)</td><td colspan="2">≤ 2 s</td><td>0.2 s to 2 s</td></tr>
<tr><td>Time duration ($\boldsymbol{t_{pd}}$) before initial depth reading ($\boldsymbol{h_i}$)</td><td colspan="2">$1\ \mathrm{s} \leq \frac{t_{pa}}{2} + t_{pd} \leq 4\ \mathrm{s}$</td><td>$\left(3 - \frac{t_{pa}}{2}\right)$ s</td></tr>
<tr><td>Additional force application time, from $\boldsymbol{F_0}$ to $\boldsymbol{F_T}(\boldsymbol{t_{aa}})$</td><td colspan="3">1 s to 8 s</td></tr>
<tr><td>Approach mean indenter velocity ($\overline{\boldsymbol{v_{fa}}}$) between $\boldsymbol{c_1 F_T}$ and $\boldsymbol{c_2 F_T}$</td><td>20 µm/s to 40 µm/s</td><td>15 µm/s to 40 µm/s</td><td>30 µm/s</td></tr>
<tr><td>Total force dwell time ($\boldsymbol{t_{td}}$)</td><td colspan="2">4 s to 6 s</td><td>5 s</td></tr>
<tr><td>Time ($\boldsymbol{t_{ar}}$) to return to $\boldsymbol{F_0}$</td><td colspan="3">≤ 2 s</td></tr>
<tr><td>Dwell time ($\boldsymbol{t_{rd}}$) for elastic recovery</td><td colspan="2">3 s to 5 s</td><td>4 s</td></tr>
<tr><td>Temperature</td><td>23 °C ± 3 °C</td><td>23 °C ± 5 °C</td><td>23 °C</td></tr>
</table>

The parameters applied in this certification are summarized in **Table 2**, followed by a direct comparison of the standards and the CCM-WGH's definition in **Table 3**. **Figure 5** illustrates the NIST Rockwell C test cycle, including the sequence of force application, dwell periods, and unloading, as defined for the SRM requirements.

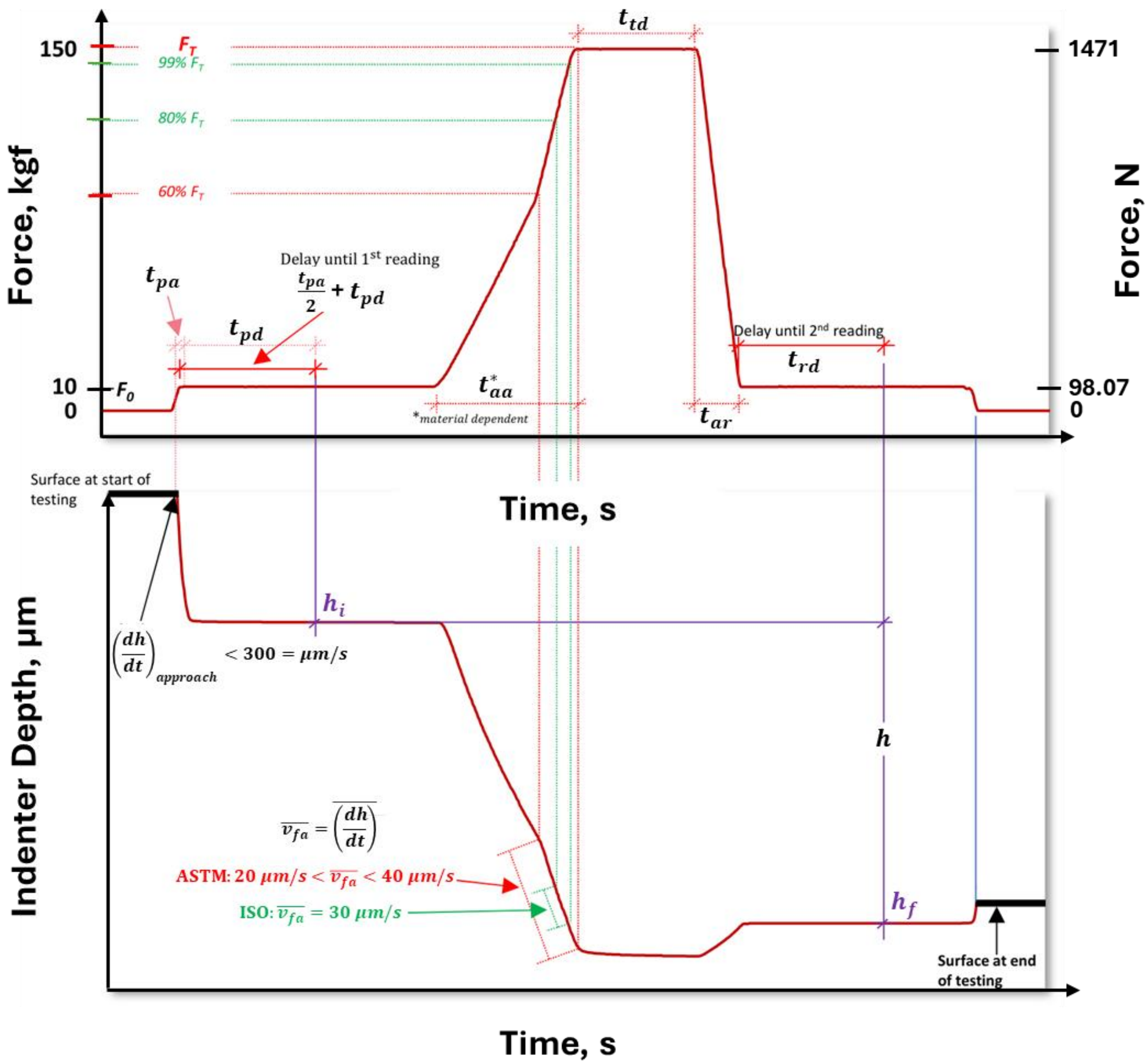


Figure 5. NIST Rockwell C test cycle plot defined for the SRM requirements. Force (top) and indenter depth (bottom) are plotted versus time.

The Rockwell C scale testing parameters established on the NIST PHSM were verified to conform to all requirements for the certification of Rockwell C hardness blocks across the three ranges (low, medium and high).

5. Direct Verifications

5.1. Verification of the Test Forces of the Primary Hardness Standardizing Machine

Prior to performing Rockwell C hardness measurements on SRMs, the test force system of the NIST PHSM was verified in accordance with [3, 5, 20]. A calibrated elastic force measuring instrument, traceable to NIST, was used to verify the force delivery throughout the force application cycle, performed without contacting a test block.

Figure 6 illustrates the force verification setup. On the left, the setup includes a force measuring instrument, a cylinder attachment positioned above it, a metal ball, and the indenter assembly. Five (5) interchangeable cylinder attachments were designed to vary the height at which the applied force was being verified. For this calibration, attachment No. 3 was used to approximate the anvil height corresponding to a NIST Rockwell C SRM block, which is nominally 15 mm thick.

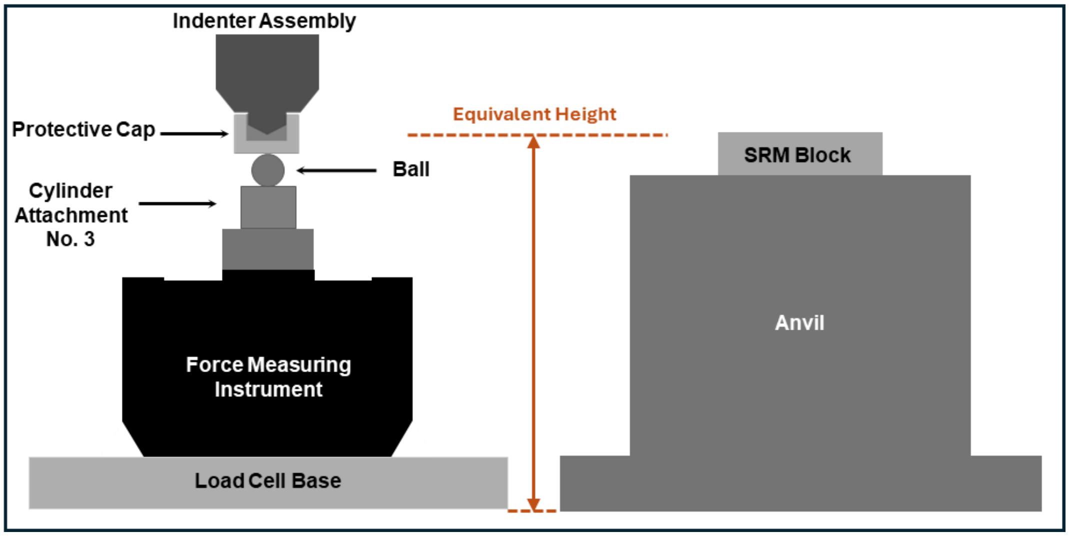


Figure 6. Illustration of the setup used to verify the test forces. Left: force measuring instrument setup including a cylinder attachment (No. 3), metal ball, and capped indenter assembly. Right: anvil of the NIST PHSM with a Rockwell C SRM block placed on top, shown for height comparison with the force verification setup.

On the right side of **Figure 6**, the NIST PHSM's anvil with an SRM block on top is shown for reference, allowing direct comparison of the setup heights. The force measuring instrument configuration (on the left) replicates the height at which the NIST Rockwell SRMs are measured; however, since the purpose of this setup is to verify force delivery rather than perform actual indentations, the indenter assembly is fitted with a protective cap to prevent contact with, or damage to, any surfaces. **Figure 7** shows a photograph of the actual force measuring instrument configuration used in the verification procedure.

Three independent calibration runs [3, 5, §5.2] were performed using the setup shown on the left side of **Figure 6**. For each run, force data were collected as a function of time. Environmental conditions, including temperature and relative humidity, were also recorded.

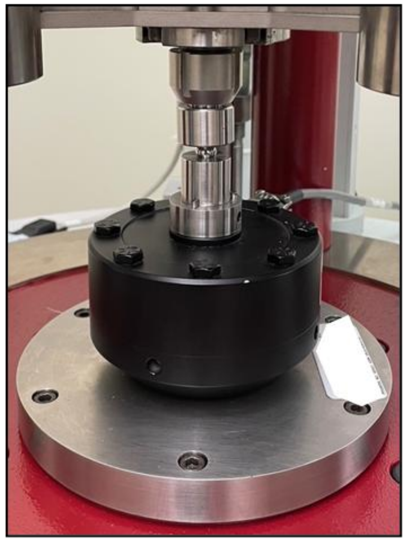


Figure 7. Actual force measuring instrument configuration used to verify the force delivery in the NIST PHSM during a force application test cycle.

To further assess the performance of the NIST PHSM, three distinct force regions [3, 5] were identified from each run, as illustrated in **Figure 8**:

1) Preliminary Force Region
2) Total Force Region
3) Recovery Region

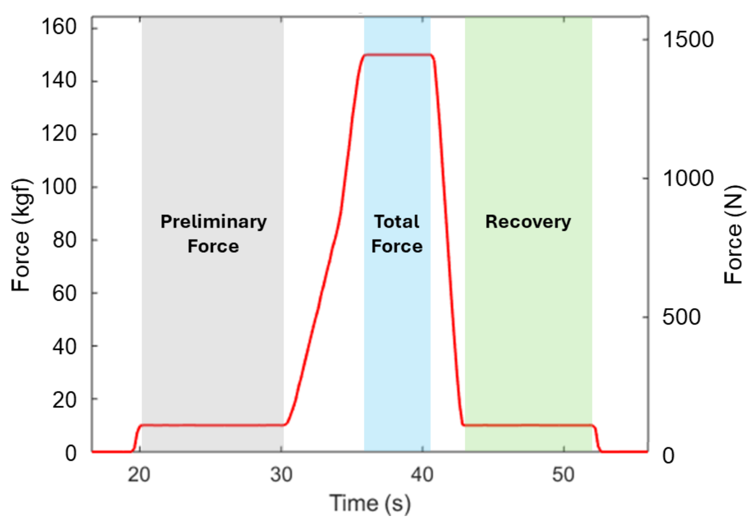


Figure 8. Force versus time diagram, with three distinct force regions evaluated during calibration runs: preliminary force, total force, and recovery.

Table 4 summarizes the mean force values recorded during three independent calibration runs and associated standard deviations using the force verification setup. The applied forces were evaluated for the three key regions of the cycle: preliminary force, total force, and recovery.

Table 4. Summary of mean forces and standard deviations for the three distinct regions: preliminary force, total force, and recovery.

	Mean Force (N)			Standard Deviation (N)		
Regions \ Runs	1	2	3	1	2	3
Preliminary Force	98.28	98.31	98.30	0.07	0.11	0.06
Total Force	1471.43	1471.34	1471.42	0.07	0.12	0.08
Recovery	98.24	98.35	98.29	0.11	0.16	0.14

The results demonstrate excellent agreement with the requirements outlined in [3, 5]. The preliminary and recovery forces across all three runs are consistently centered (within ± 0.25 %) near the nominal value of 98.07 N. All measured values are well within the ± 0.5 % tolerance specified in [3, 5]. Mean values range from 98.24 N to 98.35 N, with standard deviations between 0.06 N and 0.16 N, indicating a high degree of control and repeatability during the initial and final stages of the force cycle. The total force applied during the major force segment centers between 1471.34 N and 1471.43 N across all runs, which closely matches the nominal force requirement of 1471 N for the Rockwell C scale (i.e., 150 kgf major force). Standard deviations remain low, from 0.07 N to 0.12 N, demonstrating the machine's ability to consistently deliver stable forces throughout the full loading period.

To further highlight the consistency across the three calibration runs, the force-time profiles were aligned in **Figure 9**, such that the onset of the preliminary force region for each run overlapped. This normalized view enables direct comparison of force transitions during the preliminary, total, and recovery phases across runs.

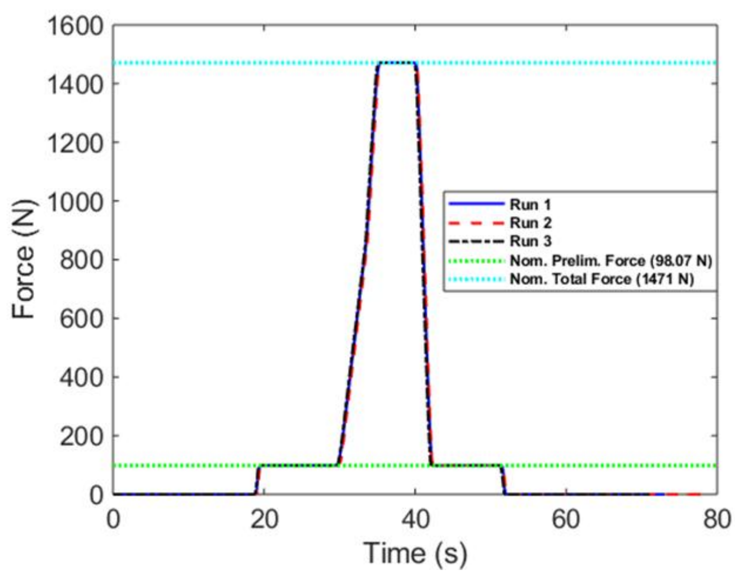


Figure 9. Aligned force versus time plots for runs 1 to 3 for direct comparison of force transitions during the preliminary, total, and recovery regions across runs. The nominal preliminary and total forces are illustrated on the figure.

The expanded uncertainty (U) of the force measuring instrument calibration was 0.0015 % at the maximum calibrated force, equivalent to ± 0.022 N at 1471 N. This uncertainty was considered when evaluating the agreement between measured forces and the values defined in [3, 5, 6]. All force measurements were performed within the Class A range of the force measuring instrument (42 N to 1468 N), except for the total force peak (1471 N), which slightly exceeded the stated maximum but remained within the effective tolerance. This confirms the force measuring instrument was suitable for both preliminary and total force evaluations in the NIST PHSM verification procedure.

These results confirm that the NIST PHSM delivered forces with high stability and repeatability throughout the entire indentation cycle, supporting its continued suitability for calibrating Rockwell C hardness reference blocks.

5.2. Depth Verification of the NIST PHSM Laser Interferometer System

The depth measuring system integrated into the NIST PHSM is based on a laser interferometer that records the vertical displacement of the indenter frame (see **Figure 1b**) and thereby of the indenter, generated by the dead-weight system. The accuracy and linearity of the interferometer system were verified using calibrated ceramic class K gauge blocks as traceable length standards. The certified blocks were stacked in different combinations to cover the full displacement range relevant to the hardness blocks tested on the NIST PHSM. The stacks were placed directly on the NIST PHSM's anvil, and the indenter was fitted with a protective cap to prevent damage to both the indenter tip and the gauge block surface. This approach mirrored the precautions adopted in the force verification procedure, and **Figure 10** illustrates the setup on the NIST PHSM's anvil. For each stack, the certified stack length and its associated uncertainty were calculated from the calibration certificates.

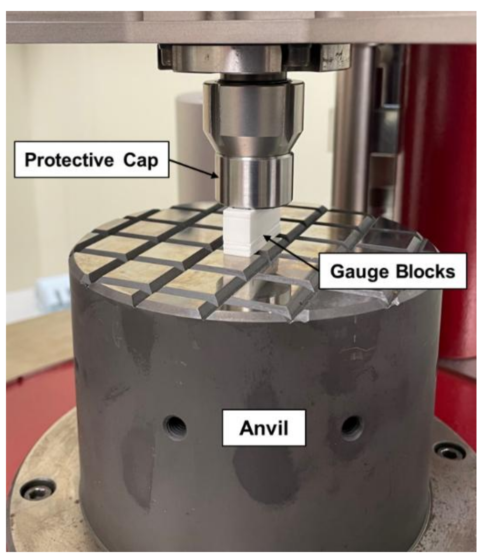


Figure 10. Illustration of the setup used for the depth verification of the NIST PHSM laser interferometer system.

5.2.1. Depth Verification Testing Process

The verification procedure consisted of:

1) Preparing the randomly selected gauge block stacks on the NIST PHSM anvil by wringing (i.e., sliding together to achieve close contact) each gauge block to the thickness of interest.
2) Protecting the indenter with a fitted cap to avoid surface damage.
3) Measuring the displacement of the indenter frame for each gauge block stack configuration ten times and recording the mean value.
4) Performing the measurements both before SRM certification testing (February 2024) and after SRM certification testing (December 2024).
5) Analyzing results both in terms of absolute stack lengths (measured versus certified) and in terms of incremental (delta) comparisons.

The incremental approach was adopted to highlight the interferometer's ability to resolve tenths of micrometer displacements while removing the influence of fixed offsets.

5.2.2. Depth Verification Results and Observations

The slope of the pre-certification results (February 2024) was 0.99989, and the slope of the post-certification results (December 2024) was 1.00003. Linear regression analysis yielded coefficients of determination (R^2) of 0.99996 and 0.99999 for the pre- and post-certification datasets, respectively, demonstrating excellent linearity over the tested displacement range. The fitted slopes, which were close in unity, indicated strong proportional agreement between the measured interferometer displacement increments and those calculated from the certified gauge block thicknesses.

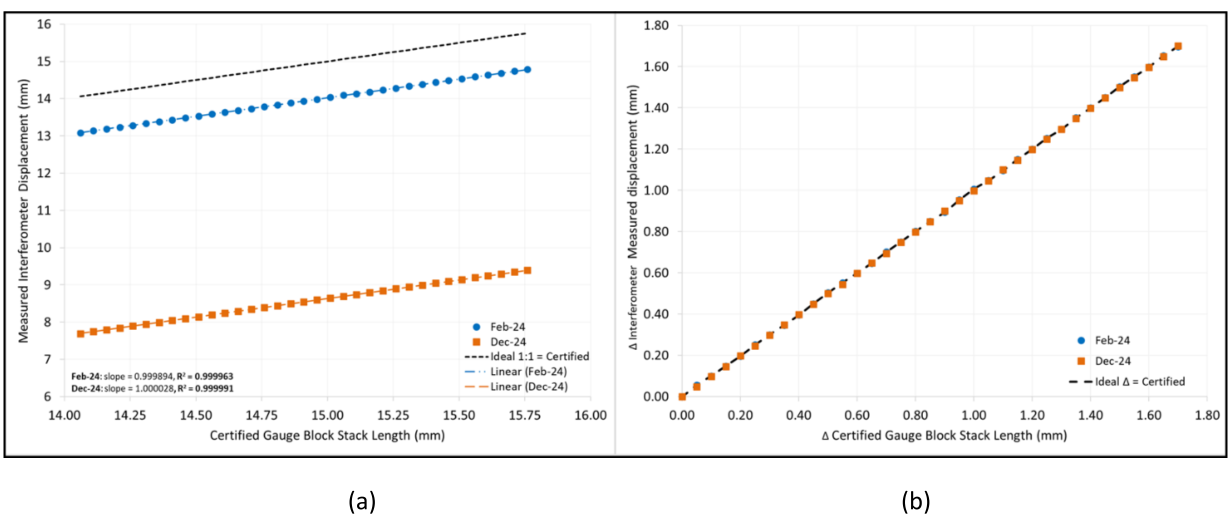


Figure 11. Interferometer verification with a) absolute displacement values compared to the certified values of the ceramic class K gauge blocks, and b) incremental differences (Δ), or comparison of measured interferometer displacement increments versus those calculated from the certified gauge block thicknesses, before and after the SRM certification.

The delta (Δ) comparison plots (**Figure 11**) present the measured interferometer displacement increments against those calculated from the certified gauge block thicknesses. For both verifications, data closely follow the ideal 1:1 line.

The pre-certification (February 2024) results show excellent agreement across the examined range, while the post-certification (December 2024) measurements demonstrated slightly improved linearity, confirming stability of the interferometer system during and after SRM certification. The incremental analysis underscores the ability of the interferometer to resolve displacement steps with high accuracy and repeatability.

The use of calibrated ceramic class K gauge blocks placed directly on the anvil, combined with indenter protection, provided a traceable and robust verification of the NIST PHSM interferometer system. The strong agreement between measured and increments determined from the certified values, observed before and after SRM testing, confirms the accuracy, stability, and traceability of the depth measuring system. This verification ensures reliable indentation depth determinations in Rockwell C hardness testing performed on the NIST PHSM.

6. Hardness Block Seating Process

In contrast to the conventional Rockwell C hardness block testing procedures, where blocks are typically hand-seated directly on the test anvil prior to indentation (performing the first HRC indentation that will not be considered in the analysis), this study employed a modified and more controlled seating process to avoid operator influence. This approach was developed to address specific challenges encountered during the certification of primary SRMs, and included enhanced control of block positioning, prevention of surface damage, and improved reproducibility of seating conditions across multiple test cycles.

The seating process described here introduces several additional components and steps beyond standard practice. These include the use of a custom-designed holder for lateral stability, a 3D-printed polymeric enclosure to guide contact to non-critical regions of the block, and a protective, non-magnetic stainless-steel cap and disk assembly to shield the indenter tip during initial setup. The overall configuration ensures that each SRM block is seated in a repeatable, damage-minimizing manner prior to the application of the Rockwell C indentation cycle.

This more elaborate method was specifically developed for the certification of NIST SRMs 2810a, 2811a, and 2812a using the NIST PHSM, where minimizing variability and maximizing control are critical to achieving the required metrological traceability.

Figure 12 illustrates the components of the seating setup. Each SRM block was positioned in a custom 3D-printed holder equipped with a positioning tray and padded side clamps (**Figure 13a**), designed to prevent lateral movement during testing. The clamps applied only minimal pressure to the sides of the block, and the bottom surface of the block was left unobstructed to ensure direct contact with the hardness testing frame's anvil. Once secured, the holder was carefully inserted into a graduated stage (**Figure 13b**) and securely fastened to the anvil, prior to initiating the seating process.

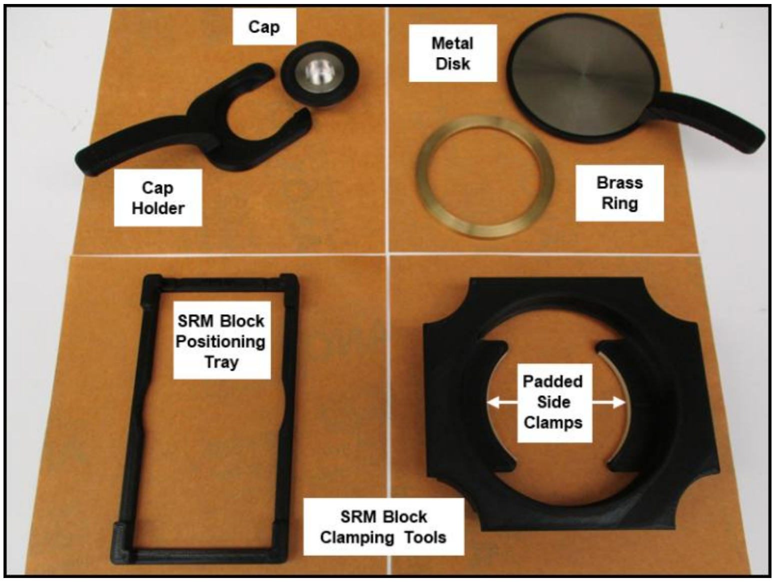


Figure 12. Components used in the SRM block seating process, including a non-magnetic cap to cover the indenter tip, a protective ring and disk assembly to safeguard the SRM block, and a positioning tray used with the clamping tool to align the block on the testing frame's anvil.

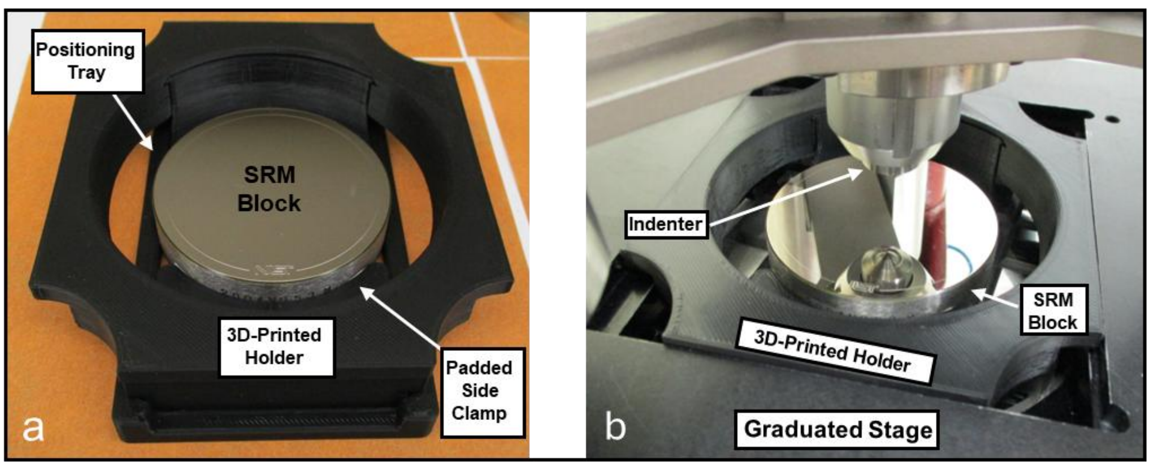


Figure 13. Sequential setup of the SRM block for testing. (a) The SRM block is secured in the custom 3D-printed holder with a block positioning tray and padded side clamps to prevent lateral movement. (b) The clamped block is inserted into the graduated stage, with the bottom surface in contact with the anvil prior to seating.

Once the holder is positioned within the graduated stage, a non-magnetic soft metal (i.e., brass) ring is placed over the SRM block, making contact only with the peripheral region (**Figure 14a**). Brass was selected to minimize the risk of surface damage, even though the region outside the engraved circle is not used for hardness measurements. A stainless steel disk, enclosed within a 3D-printed polymeric holder, is then placed on top of the brass ring. The enclosure helps prevent unwanted scratches on the block's surface, and the stainless steel disk contacts only the brass ring on the peripheral region of the block, which is outside the primary test area defined by the circular engravement (**Figure 14b**). The polymeric material used for the enclosure is the same as that used for the SRM block holder. Finally, a non-magnetic cap is placed over the stainless steel disk to shield the indenter tip and prevent unintended contact with any surface during the seating process (**Figure 14b** illustrates the seating setup on the testing machine).

The seating process consisted of positioning the stainless steel cap and applying the total test force of 1471 N (150 kgf) to the SRM block for 15 s, followed by removal of the force. This step was repeated once more before removing the cap and disk. At that point, the block was considered properly seated on the testing frame's anvil and ready for the first indentation.

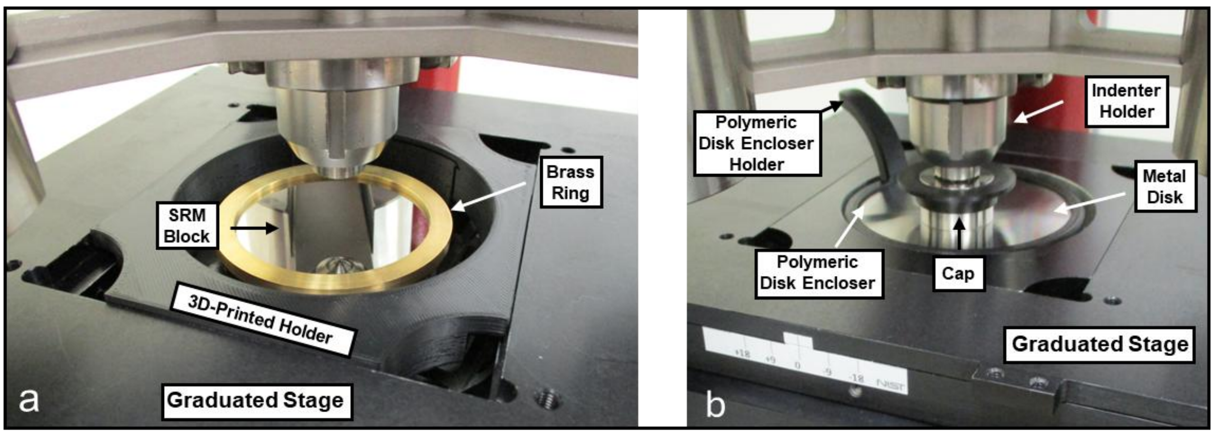


Figure 14. Components used for the seating process on the NIST PHSM, including (a) the brass ring, and (b) the stainless steel disk and non-magnetic cap for indenter protection.

7. Indentation Testing Locations

A total of thirteen (13) indentations were performed on each SRM block using predetermined X and Y indentation locations arranged in a fixed radially symmetric pattern. The order in which the thirteen locations were indented was randomized for each block. **Table 5** provides an example of two medium-range SRM blocks, each indented thirteen times at the predefined X and Y locations. Although the indentation sequence was randomized, the overall radially symmetric pattern, which resembles a star, was preserved for each tested block (**Figure 15**).

Table 5. Examples of X and Y coordinates of the thirteen (13) indentations that were performed on each SRM block in a randomized sequence.

	SRM Block Number			
	B692		B693	
Indentation Location and Number	X (in mm)	Y (in mm)	X (in mm)	Y (in mm)
1	18	-10.4	0	10.4
2	0	-10.4	9	5.2
3	9	-5.2	9	-5.2
4	-18	-10.4	-18	-10.4
5	-18	10.4	0	20.8
6	18	10.4	-9	5.2
7	0	20.8	0	-20.8
8	-9	-5.2	0	-10.4
9	0	10.4	18	10.4
10	-9	5.2	-9	-5.2
11	0	-20.8	-18	10.4
12	0	0	18	-10.4
13	9	5.2	0	0

Following each indentation, the block was repositioned using the graduated stage (**Figure 14b**). Because repositioning disrupted the contact between the block and the anvil, the seating procedure was repeated to ensure proper contact between the block and the anvil before each indentation. All hardness values resulting from the thirteen (13) indentations were included in the measurements to determine the SRM certified hardness values for that SRM.

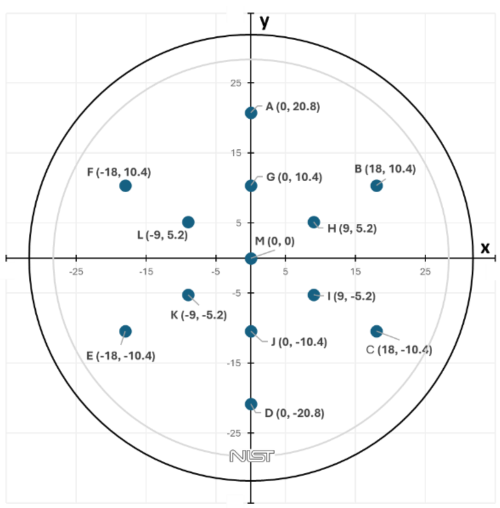


Figure 15. Test surface of an SRM block showing the locations of the certified indentations (dots), arranged in a radially symmetric pattern. Indentations were performed in a randomized sequence for each SRM. The engraved circle and NIST logo are also illustrated. The virtual axes and coordinates shown on the block surface are in millimeters (mm), with the block center located at M (0, 0).

8. Verification of Traceability Using Previously Certified SRMs

To maintain continuity of traceability from the previous NIST hardness standardizing machine to the NIST PHSM, previously certified SRMs were retested using the NIST master standard Rockwell diamond indenter (No. 3581) [21]. Two previously certified blocks were selected for evaluation (**Table 6**):

- SRM 2811, Block 97C45601, nominal hardness 45 HRC
- SRM 2812, Block 97C63202, nominal hardness 63 HRC

These blocks were originally certified on the previous generation hardness standardizing machine and subsequently remeasured on the NIST PHSM. A block of nominal hardness 25 HRC was not retested due to the lack of available inventory.

Table 6. Result summary of previously certified SRMs retested using the NIST master standard Rockwell diamond indenter (No. 3581) on the NIST PHSM. The nonuniformity is the difference between the highest to the lowest hardness value [3, §A4.5.4, 5, §8.1].

Block No.	Nominal Hardness	Year	Average Hardness (HRC)	Standard Deviation, σ (HRC)	Nonuniformity (HRC)
97C45601	45 HRC	2009	45.47	0.04	0.12
		2024	45.62	0.01	0.02
97C63202	63 HRC	2009	64.11	0.07	0.20
		2024	64.22	0.04	0.10

- For the 45 HRC Block (97C45601, SRM 2811):

Seven (7) measurements obtained in 2009 yielded an average hardness of 45.47 HRC with a standard deviation of 0.04 HRC. In 2024, three (3) measurements performed on the NIST PHSM yielded an average hardness of 45.62 HRC, corresponding to a difference of 0.15 HRC. The NIST PHSM measurements demonstrated lower variability (σ = 0.01 HRC versus 0.04 HRC) and lower nonuniformity (0.02 HRC versus 0.12 HRC). A Welch two-sample t-test comparing the 2009 and 2024 datasets yielded a statistically significant difference in the mean hardness values (t = 8.00, p < 0.001), where t is the test statistic quantifying the normalized difference between the two mean values and p is the probability of observing a difference at least this large if no true difference exists between the datasets.

- For the 63 HRC Block (97C63202, SRM 2812):

Seven (7) measurements obtained in 2009 yielded an average of 64.11 HRC, with a standard deviation of 0.07 HRC. In 2024, four (4) measurements performed on the NIST PHSM yielded an average hardness of 64.22 HRC, corresponding to a difference of 0.11 HRC. The NIST PHSM measurements again exhibited lower variability (σ = 0.04 HRC versus 0.07 HRC) and lower nonuniformity (0.10 HRC versus 0.20 HRC). A Welch two-sample t-test similarly indicated a statistically significant difference between the earlier and later measurements (t = 3.26, p = 0.01).

Although statistically significant differences were observed between the historical certification measurements and the NIST PHSM retest measurements, the absolute differences were small (≤ 0.15 HRC). In addition, the measured nonuniformity values for all measurements remained below the maximum permissible limits for HRC standardized test blocks specified in ASTM [3, §Table A4.2] and ISO [5, §Table 5]. Together with the lower measurement variability and nonuniformity observed on the NIST PHSM, these results support continuity of hardness realization and traceability between the previous NIST hardness standardizing machine and the NIST PHSM. Therefore, although statistical differences were detected, the observed differences are considered metrologically small and do not indicate a meaningful change in the realization of the Rockwell C hardness scale.

Furthermore, intra-laboratory comparison exercises were conducted on the former NIST PHSM using the same NIST master indenter (No. 3581) [21]. Since this indenter is now used on the current NIST PHSM for certifying the new batch of Rockwell C SRMs (2810a, 2811a, and 2812a), the traceability chain is directly linked to internationally validated results.

9. Statistical Analysis and Estimation Methods

With the experimental methods fully established, the next portion of this report addresses the reduction of the experimental data, and the statistical methodology applied to support certification. The goal is to transform the raw hardness determinations into certified values for SRMs 2810a, 2811a, and 2812a, while providing uncertainties that accurately reflect both the measurement process and the needs of the end users. This section begins by outlining the rationale for the selected approaches, then introduces two alternative certification modalities: 1) block-based (B) and 2) set-based (S). Consistent with international standards and in alignment with what users expect, NIST adopted only the block-based modality (B) for the 2024 certification of these SRMs. Nevertheless, the set-based modality (S) is also presented, because its evaluation contributes additional evidence supporting the robustness and overall quality of the certified SRM blocks.

The statistical methodology applied in this certification effort was designed to ensure that the reported values and uncertainties are both scientifically defensible and practically useful to the community of metallic materials Rockwell C hardness users. The analysis framework relies on robust estimation procedures and consensus-building models that account for the characteristics of hardness testing data, including its sensitivity to occasional abnormal values and the need for repeatability assessments that are not overly influenced by such occurrences.

9.1. Modalities

9.1.1. Block-Based Certification (Modality B)

In modality B, each block receives a block-specific certified value, estimated from the thirteen (13) individual determinations made on the same test surface. Rather than adopting the conventional (equally weighted) arithmetic average of these determinations, as prescribed in ASTM [3, §3.2.1], the NIST analysis proposed estimates of average hardness that are a robust version of the arithmetic average described in **section 9.2.1. on Average Hardness**. These estimates achieve nearly the same efficiency as the arithmetic mean when all values exhibit approximately symmetric behavior without substantial anomalous observations, while offering strong resistance to the occasional presence of abnormal determinations that cannot be excluded on substantive grounds. This approach follows the principle articulated by the Analytical Methods Committee of the Royal Society of Chemistry [22], which cautions against rejecting “outliers” without a defensible cause.

In evaluating repeatability as a source of uncertainty surrounding measured values of hardness, the analysis also departs from the conventional approach of ASTM [3, §3.2.3]. Here also, a robust way to quantify the dispersion of replicate determinations of hardness made on each block under conditions of repeatability – described in **section 9.2.2.** on **Block-Specific Uncertainty Evaluations** − is adopted.

9.1.2. Set-Based Certification (Modality S)

In modality S, the certified value assigned to an SRM lot is common to all its blocks. It is obtained as a consensus estimate of the block-specific averages determined in modality B. This consensus calculation incorporates both the uncertainties and associated degrees of freedom of the block-specific averages, as well as the numbers of degrees of freedom that support their evaluations.

A consensus value is a particular weighted average of block-specific averages, with data-driven weights that take into account the uncertainties associated with the block-specific averages and any “excess” dispersion of the block-specific averages. “Excess” means above and beyond the dispersion suggested by the standard uncertainties associated with the block-specific averages. Consensus values were computed separately for SRMs 2810a, 2811a, and 2812a, using the state of-the-art consensus building models and methods implemented in the NIST Decision Tree [23].

Taken together, these approaches yield certified values that balance statistical rigor with practical robustness.

For each SRM lot, candidate values and associated uncertainties for two certification modalities are presented in this report, with two variants per modality (A for average and P for predicted):

In **modality B**, each block has its own certified average value of hardness, $\overline{H}$, and associated expanded uncertainty, $U_{95}(\overline{H})$, corresponding to 95 % coverage. In addition, a corresponding prediction uncertainty, $U_{95,P}(H)$, is listed.

> **BA –** NIST believes that the block’s true average (A) hardness lies in the interval $\overline{H} \pm U_{95}(\overline{H})$ with 95 % probability.
>
> **BP –** NIST believes that the predicted (P) hardness value that will be measured at an indentation made on the block’s surface, at a location sufficiently far apart from any

preexisting indentation, and using NIST's test method and a Rockwell hardness machine comparable to the NIST PHSM used during certification, will lie in the interval $\overline{H} \pm U_{95,P}(H)$ with 95 % probability.

In **modality S**, if the SRM lot is sufficiently homogeneous, the same certified hardness value, $\widetilde{H}$, can be assigned to all blocks in the lot. This value represents a consensus of the block-specific average hardness values and is qualified by the expanded uncertainties $U_{95}(\widetilde{H})$ and $U_{95,P}(H)$, with the following meaning:

SA – NIST believes that the average hardness of all the blocks in the SRM lot lies in the interval $\widetilde{H} \pm U_{95}(\widetilde{H})$ with 95 % probability.

SP – NIST believes that the hardness value that will be measured at an indentation made on the surface of any block in the SRM lot, at a location sufficiently far apart from any preexisting indentation on this block, and using NIST's test method and a Rockwell hardness machine comparable to the NIST PHSM used during certification, will lie in the interval $\widetilde{H} \pm U_{95,P}(H)$ with 95 % probability.

It is important to note that only modality B, including both variants BA and BP, is on the certificate of each SRM 2810a, 2811a, 2812a. Modality S is only presented in this report.

9.2. Statistical Models and Methods

The following thirteen (13) determinations of the hardness of block 711 measured from SRM 2812a: 62.71, 62.73, 62.71, 62.71, 62.69, 62.69, 62.71, 62.71, 62.72, 62.70, 62.66, 62.58, 62.70 in unit of HRC are considered. The conventional estimation of the average hardness involves the assumption that they are like a sample from a Gaussian distribution. If such an assumption were valid, then the best estimate of the average hardness would be their arithmetic average, $\overline{H}$ = 62.69 HRC, and the expanded uncertainty for 95 % coverage would be half the length of the Student's t 95 % confidence interval for the true average, based on 13−1 = 12 degrees of freedom, $U_{95}(\overline{H})$ = 0.0461.

However, the Shapiro-Wilk test [24] rejects unequivocally (p-value of 0.0005) the assumption of Gaussian shape for this set of determinations (**Figure 16**), and none of the results listed in the previous paragraph are valid.

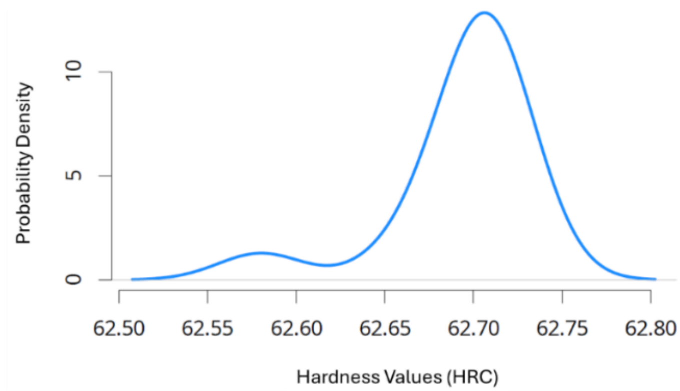


Figure 16. Probability density estimate of Rockwell C hardness measured on the surface of block B711 (of SRM 2812a) under conditions of repeatability.

9.2.1. Average Hardness

The Hodges-Lehmann (HL) estimator was adopted to compute the average hardness of each block because it provides considerable robustness and efficiency, and both the standard and expanded uncertainties can easily be computed by inversion of Wilcoxon's signed-rank test [25, 26].

The HL estimator is immune to the presence of up to 29 % of outliers (which translates to 4 outliers in a set of 13 determinations), and its variance is only 5 % larger than the variance of the arithmetic average when it is used to estimate the average of a Gaussian sample.

However, for samples from distributions whose tails are heavier than Gaussian tails, the HL estimator can be much more efficient than the arithmetic average. For example, for samples of size 13 from a Student's t distribution with 3 degrees of freedom, the standard error of the HL estimates of the true average is approximately 1.3 times smaller than the standard error of the arithmetic average.

If the underlying distribution of the hardness determinations made on the same block is approximately symmetrical, then the HL estimator is an unbiased estimator of the average (and median) hardness. In other cases, it estimates a "pseudo-median" that is close to the median.

To compute the HL estimate for the set of 13 determinations listed above, the following steps can be taken, or the function wilcox.test can simply be invoked and is defined in the R environment for statistical programming and graphics [27]:

For **HL.1**, form a set whose elements are the 13 determinations and the arithmetic averages of the $13 * \frac{(13-1)}{2} = 78$ subsets of size 2 of the set of 13 determinations: 62.71, ..., 62.70, $\frac{(62.71+62.73)}{2}$, ..., $\frac{(62.66+62.58)}{2}$. These are the 13 + 78 = 91 *Walsh averages* that correspond to the 13 determinations under consideration.

For **HL.2**, compute the median of these *Walsh averages*.

For a Gaussian distribution with mean μ and standard deviation σ, the interval from $\mu - \sigma$ to $\mu + \sigma$ includes 68 % of the distribution's unit of probability. Considering this, a proxy can be defined for the standard uncertainty associated with an estimate m of the mean μ of a symmetrical, but not necessarily Gaussian, distribution as $u(m)$, such that $m \pm u(m)$ is a 68 % coverage interval for μ [28]. This is the motivation for the procedure proposed by Stoudt et al. [26] to evaluate the standard uncertainty of the HL estimates, based on samples as small as the blocks in the SRMs under consideration. For larger samples, the nonparametric bootstrap can also be used for this purpose.

Listing 1 presents an R implementation of the procedure used to estimate the average hardness of a block, and for evaluating the standard and expanded uncertainties associated with this average and the expanded uncertainty for prediction, as applied to the determinations made on block 711 from SRM 2812a for purposes of illustration.

```
x = c(62.71, 62.73, 62.71, 62.71, 62.69, 62.69, 62.71,
62.71, 62.72, 62.70, 62.66, 62.58, 62.70)
w68 = wilcox.test(x, conf.int=TRUE, conf=pnorm(1)-pnorm(-1))
HL = w68$estimate; names(HL) = NULL
uHL = diff(w68$conf.int)/2; names(uHL) = NULL
c(HL=HL, uHL=uHL) ## 62.705 0.0075

uDD.2812 = 0.036; uST.2812 = 0.020; uSI.2812 = 0.067
uc = sqrt(uHL^2 + uDD.2812^2 + uST.2812^2 + uSI.2812^2); uc ## 0.079
U95.A = qt(0.975, df=12) * uc;                    U95.A ## 0.17
U95.P = qt(0.975, df=12) * uc * sqrt(13 + 1); U95.P ## 0.64
```

Listing 1. R code implementing the HL estimator and the corresponding uncertainty evaluations (Sections 9.2.1 and 9.2.2), applied to the replicated hardness measurements of block 711 from SRM 2812a, and corresponding to the uncertainty budget for this SRM listed in Table 7. The uncertainty component, u_A, in Equation (3) and Table 7 is the uHL in the listing. The expanded uncertainty associated with the block's average hardness is U95.A, and U95.P is its counterpart for predictions of future measurements of hardness made at individual locations sufficiently far away from any preexisting indentations.

9.2.2. Block-Specific Uncertainty Evaluations

Table 7 lists the uncertainty budgets for specific blocks in SRMs 2810a, 2811a, and 2812a, for illustrative purposes.

The uncertainty budgets for the other blocks in the corresponding SRMs are close to those listed in this table; only the values of the material uniformity and measurement repeatability, u_A, the combined standard uncertainty, $u_c(\overline{H})$, the expanded uncertainties for 95 % coverage, $U_{95}(\overline{H})$ (for the average hardness, $\overline{H}$) and of $U_{95,P}(H)$ for a prediction of the hardness value, H, that will be measured at a future indentation, vary, albeit only slightly, from block to block. The contributions from the other sources listed are the same that have been used in previous lots of Rockwell C SRMs, issued on March 20th, 2013.

Table 7. The illustrative uncertainty budgets for specific blocks of SRMs 2810a (Block 723), 2811a (Block 674), and 2812a (Block 696) are listed. Only the illustrative uncertainty contributions listed for u_A, $u_c(\overline{H})$, $U_{95}(\overline{H})$ (for the average hardness, $\overline{H}$), and $U_{95,P}(H)$ (for a prediction of the hardness value, *H*, that will be measured at a future indentation) are block-specific, with the values listed being the medians of the evaluations of the block-specific contributions from these sources of uncertainty.

SRM 2810a Low Rockwell C (Block 723)		
u_{A}	Material Uniformity and Measurement Repeatability	0.017
u_{DD}	Day-to-Day Reproducibility	0.080
u_{ST}	NIST Hardness Standardizing Machine	0.007
u_{SI}	NIST Standardizing Indenter	0.067
$u_{\mathrm{c}}(\overline{H})$	Combined Standard Uncertainty	0.11
$U_{95}(\overline{H})$	Expanded Uncertainty (95 % Coverage) for $\overline{H}$	0.22
$U_{95,P}(H)$	Expanded Uncertainty (95 % Coverage) for prediction	0.86

SRM 2811a Mid Rockwell C (Block 674)		
u_{A}	Material Uniformity and Measurement Repeatability	0.015
u_{DD}	Day-to-Day Reproducibility	0.038
u_{ST}	NIST Hardness Standardizing Machine	0.007
u_{SI}	NIST Standardizing Indenter	0.067
$u_{\mathrm{c}}(\overline{H})$	Combined Standard Uncertainty	0.079
$U_{95}(\overline{H})$	Expanded Uncertainty (95% Coverage) for $\overline{H}$	0.17
$U_{95,P}(H)$	Expanded Uncertainty (95% Coverage) for prediction	0.64

SRM 2812a High Rockwell C (Block 696)		
u_{A}	Material Uniformity and Measurement Repeatability	0.010
u_{DD}	Day-to-Day Reproducibility	0.023
u_{ST}	NIST Hardness Standardizing Machine	0.007
u_{SI}	NIST Standardizing Indenter	0.067
$u_{\mathrm{c}}(\overline{H})$	Combined Standard Uncertainty	0.072
$U_{95}(\overline{H})$	Expanded Uncertainty (95% Coverage) for $\overline{H}$	0.15
$U_{95,P}(H)$	Expanded Uncertainty (95% Coverage) for prediction	0.59

The block-specific combined standard uncertainties and expanded uncertainties were computed as follows [29, §4.2, §4.7] (u_A typically varies from block to block):

$$u_c(\overline{H}) = (u_A^2 + u_{DD}^2 + u_{ST}^2 + u_{SI}^2)^{½} \tag{3}$$

$$U_{95}(\overline{H}) = Q_t(0.975, 12) \times u_c(\overline{H}) \tag{4}$$

$$U_{95,P}(H) = Q_t(0.975, 12) \times u_c(\overline{H})\sqrt{13 + 1}, \tag{5}$$

where $Q_t(0.975, 12)$ denotes the 97.5th percentile of the Student's t distribution with 12 degrees of freedom (considering that 13 indentations were made on each block for purposes of certification).

The use of $Q_t(0.975, 12)$ as the coverage factor is a rough approximation because it hinges on the assumption that $(\overline{H} - H) / u(\overline{H})$ has a Student's t distribution with 12 degrees of freedom, where H denotes the block's true average hardness, and $u(\overline{H})$ is evaluated as a proxy for the standard error of the Hodges-Lehmann estimator, by inversion of Wilcoxon's signed-rank test, as illustrated in **Listing 1**.

It should be noted that the expanded uncertainties in **Equations (4)** and **(5)** are only approximations of their true values because they assume that $\overline{H}$ is the average of a Gaussian sample, while in fact it is the Hodges-Lehmann estimate, which is approximately Gaussian when it is applied to large samples (from possibly non-Gaussian distributions).

Some exploratory simulations suggest that Student's t can be a fairly accurate approximation to the distribution of the ratio $(\overline{H} - H) / u(\overline{H})$, but with fewer than the hypothesized 12 degrees of freedom (more like 9 degrees of freedom).

The expanded uncertainty, $U_{95}(\overline{H})$ in **Equation (4)**, that qualifies a block's certified value, $\overline{H}$, expresses NIST's belief that the true value of the block's average hardness lies in the interval $\overline{H} \pm U_{95}(\overline{H})$ with 95 % probability.

The prediction of the hardness value that will be measured at a new test location spatially independent [3, §7.9, 4, §8.13], such that the plastic deformation fields do not overlap with any preexisting indentations, is $\overline{H}$, but the expanded uncertainty associated with the prediction, $U_{95,P}(H)$ in **Equation (5)**, is appreciably larger than $U_{95}(\overline{H})$. The corresponding formula is a particular case of Equation (4.7) in Meeker et al. [29].

9.2.3. SRM-Specific Averages and Uncertainty Evaluations

Figure 17 shows that most intervals of the form, $\overline{H}_j \pm U_{95}(\overline{H}_j)$, overlap considerably, indicating that the measurement results of the block-specific average values are generally consistent with one another. This suggests that it may be acceptable to assign the same, single-certified value to all the blocks in the SRM lot. However, substantial overlap of the intervals is not a strict requirement for assigning a common certified value. A single value may still be justified even when differences among block-specific averages are larger than expected from their individual uncertainties, provided that the overall variability among blocks remains acceptable under the consensus model described below.

To explore this possibility, models of the following form were fitted to the sets of values of block-specific average hardness for each of the three SRMs under consideration, with the results reported in **Section 9.3.2.** on **SRM-Specific Certification (Modality S)**:

$$\overline{H}_j = H + \beta_j + \varepsilon_j, \text{ for } j = 1, \ldots, n, \qquad \textbf{(6)}$$

where H denotes the consensus value of the true values of the block-specific averages, β_j a block effect indicating whether $\overline{H}_j$ is estimated to be above or below H, and ε_j the measurement error specific to the average hardness of block j. Also, in **Section 9.3.2.**, $\widetilde{H}$ will be denoted by the estimate of the consensus value H in **Equation (6)**, to distinguish it from the block-specific hardness averages $\{\overline{H}_j\}$.

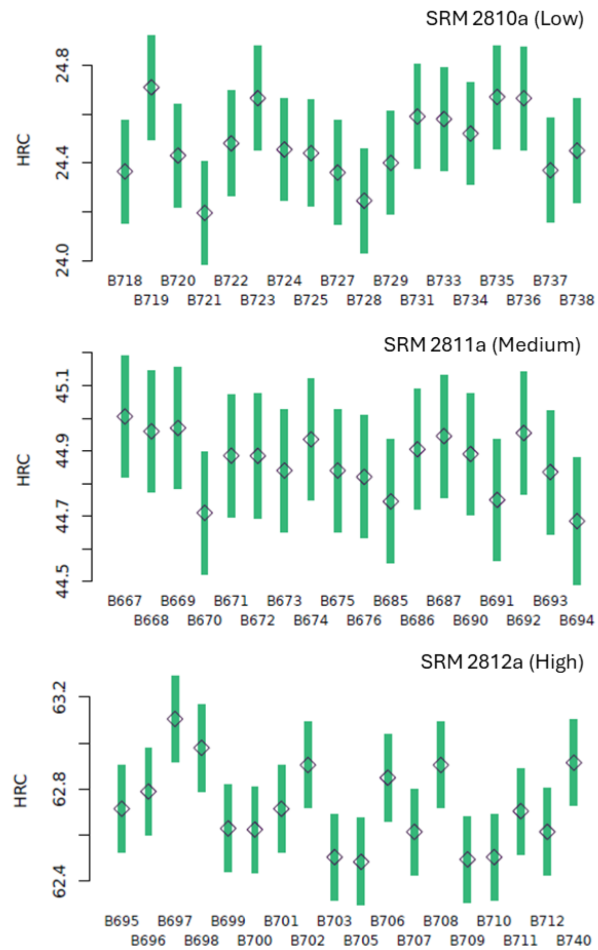


Figure 17. Block-specific estimates of the average hardness $\overline{H}_1, \ldots, \overline{H}_n$ of the n =18 blocks in each SRM are indicated by diamonds. The thick (green) vertical line segments represent $\overline{H}_j \pm U_{95}(\overline{H}_j)$ for j = 1, …, n.

The NIST Decision Tree [23] recommended the Adaptive Weighted Average (AWA) as the consensus value for the block averages of SRM 2811a, and the hierarchical (Bayesian) model labeled with Gaussian block effects and Gaussian measurement errors (often referred to as HGG, where H means "hierarchical", the first G indicates that the laboratory effects are Gaussian, and the second G indicates that the measurement errors are Gaussian) for SRMs 2810a and 2812a. Since AWA is a special case of HGG, the same HGG model is fitted (but with different sets of parameter values) to the block averages of all three SRMs: the block effects $\{\beta_j\}$ are similar to a sample from a Gaussian distribution with mean 0 HRC and standard deviation τ, and the measurement errors $\{\varepsilon_j\}$ are like outcomes of independent Gaussian random variables, all with mean 0 HRC, but possibly different standard deviations $\{\sigma_j\}$.

The standard deviation τ quantifies the so-called *dark uncertainty*, which expresses the "excess" dispersion of the block averages above and beyond what their associated uncertainties suggest this dispersion should be. When τ does not differ significantly from 0 HRC, the block averages are considered mutually consistent.

9.3. Certification

9.3.1. Block-Specific Certification (Modality B)

Figure 17 depicts the values of average hardness of the individual blocks for each of the three SRMs under consideration, and their associated expanded uncertainties. These blocks' average hardness, $\overline{H}$, the contribution from (or lack of) material uniformity and measurement repeatability, u_A, their associated combined standard uncertainty, $u_c(\overline{H})$, and the expanded uncertainty, $U_{95}(\overline{H})$, are listed in **Tables 8** to **10**. The expanded uncertainty, $U_{95}(\overline{H})$, associated with $\overline{H}$ such that $\overline{H} \pm U_{95}(\overline{H})$ is believed to include the true value of the block's average hardness with a 95 % probability. The expanded prediction uncertainty, $U_{95,P}(H)$, is such that a future measurement of hardness made on the block at a location sufficiently far away from preexisting indentations, using NIST's test method and a Rockwell hardness machine comparable to the NIST PHSM used for the certification of the material, is expected to yield a value in the interval $\overline{H} \pm U_{95,P}(H)$ with 95 % probability.

The same tables also list the expanded uncertainties for predictions of future measurements of hardness at individual locations on each block, $U_{95,P}(H)$. The value of hardness predicted at each such indentation is the block's average, but the corresponding uncertainty is much larger than the expanded uncertainty associated with the block's average, as expected.

Table 8. SRM 2810a: Block average hardness, $\overline{H}$, material uniformity and measurement repeatability, u_A, with associated combined standard uncertainty, $u_c(\overline{H})$, expanded uncertainty, $U_{95}(\overline{H})$, and expanded prediction uncertainty, $U_{95,P}(H)$ are listed.

	$\overline{H}$	u_A	$u_c(\overline{H})$	$U_{95}(\overline{H})$	$U_{95,P}(H)$
Block	/ HRC				
B718	24.36	0.010	0.09	0.19	0.73
B719	24.71	0.020	0.09	0.19	0.75
B720	24.43	0.012	0.09	0.19	0.74
B721	24.19	0.015	0.09	0.19	0.74
B722	24.48	0.025	0.09	0.20	0.76
B723	24.67	0.017	0.09	0.19	0.74
B724	24.45	0.010	0.09	0.19	0.74
B725	24.44	0.027	0.09	0.20	0.76
B727	24.36	0.018	0.09	0.19	0.74
B728	24.24	0.017	0.09	0.19	0.74
B729	24.40	0.015	0.09	0.19	0.74
B731	24.59	0.020	0.09	0.19	0.75
B733	24.58	0.017	0.09	0.19	0.74
B734	24.52	0.010	0.09	0.19	0.73
B735	24.67	0.013	0.09	0.19	0.74
B736	24.66	0.015	0.09	0.19	0.74
B737	24.37	0.020	0.09	0.19	0.75
B738	24.45	0.020	0.09	0.19	0.75

Table 9. SRM 2811a: Block average hardness, $\overline{H}$, material uniformity and measurement repeatability, u_A, with associated combined standard uncertainty, $u_c(\overline{H})$, expanded uncertainty, $U_{95}(\overline{H})$, and expanded prediction uncertainty, $U_{95,P}(H)$ are listed.

	$\overline{H}$	u_A	$u_c(\overline{H})$	$U_{95}(\overline{H})$	$U_{95,P}(H)$
Block	/ HRC				
B667	45.00	0.015	0.08	0.17	0.66
B668	44.96	0.012	0.08	0.17	0.66
B669	44.97	0.013	0.08	0.17	0.66
B670	44.71	0.015	0.08	0.17	0.66
B671	44.89	0.018	0.08	0.17	0.66
B672	44.88	0.025	0.08	0.18	0.68
B673	44.84	0.015	0.08	0.17	0.66
B674	44.93	0.015	0.08	0.17	0.66
B675	44.84	0.015	0.08	0.17	0.66
B676	44.82	0.017	0.08	0.17	0.66
B685	44.75	0.020	0.08	0.17	0.67
B686	44.90	0.005	0.08	0.17	0.65
B687	44.95	0.017	0.08	0.17	0.66
B690	44.89	0.013	0.08	0.17	0.66
B691	44.75	0.015	0.08	0.17	0.66
B692	44.95	0.018	0.08	0.17	0.66
B693	44.84	0.020	0.08	0.17	0.67
B694	44.68	0.030	0.08	0.18	0.69

Table 10. SRM 2812a: Block average hardness, $\overline{H}$, material uniformity and measurement repeatability, u_A, with associated combined standard uncertainty, $u_c(\overline{H})$, expanded uncertainty, $U_{95}(\overline{H})$, and expanded prediction uncertainty, $U_{95,P}(H)$ are listed.

	$\overline{H}$	u_A	$u_c(\overline{H})$	$U_{95}(\overline{H})$	$U_{95,P}(H)$
Block			/ HRC		
B695	62.71	0.010	0.08	0.17	0.65
B696	62.79	0.010	0.08	0.17	0.65
B697	63.10	0.008	0.08	0.17	0.64
B698	62.98	0.015	0.08	0.17	0.65
B699	62.63	0.013	0.08	0.17	0.65
B700	62.62	0.008	0.08	0.17	0.64
B701	62.71	0.010	0.08	0.17	0.65
B702	62.91	0.010	0.08	0.17	0.65
B703	62.50	0.008	0.08	0.17	0.64
B705	62.48	0.015	0.08	0.17	0.65
B706	62.85	0.010	0.08	0.17	0.65
B707	62.62	0.010	0.08	0.17	0.65
B708	62.90	0.010	0.08	0.17	0.65
B709	62.49	0.010	0.08	0.17	0.65
B710	62.51	0.012	0.08	0.17	0.65
B711	62.70	0.007	0.08	0.17	0.64
B712	62.62	0.010	0.08	0.17	0.65
B740	62.91	0.010	0.08	0.17	0.65

9.3.2. SRM-Specific Certification (Modality S)

As already noted in **Section 9.2.3.** on **SRM-Specific Averages and Uncertainty Evaluations**, **Figure 17** also shows that the ranges of most of the intervals of the form $\overline{H}_j + U_{95}(\overline{H}_j)$ overlap considerably, suggesting that it may be acceptable to assign the same, single certified value to all the blocks in the SRM.

For each of SRMs 2810a, 2811a, and 2812a, **Table 11** lists the consensus value, $\widetilde{H}$, derived from the block-specific values of average hardness, its associated standard uncertainty, $u(\widetilde{H})$, and expanded uncertainty, $U_{95}(\widetilde{H})$, as well as the estimate of τ, which quantifies dark uncertainty.

As already pointed out in **Section 9.2.3.**, the statistical model fitted to the block-specific averages using the NIST Decision Tree, was the hierarchical (Bayesian) model with Gaussian block effects and Gaussian measurement errors (HGG). The consensus values are indicated by the horizontal lines in **Figure 18**.

Table 11. Characterization of the three SRMs produced by the NIST Decision Tree when all blocks in the same SRM are assigned the consensus value, $\widetilde{H}$.

		$\widetilde{H}$	$u(\widetilde{H})$	$U_{95}(\widetilde{H})$	$\tilde{\tau}$	$U_{95,P}(H)$
SRM	**Model**	**/ HRC**				
2810a	HGG	24.48	0.032	0.068	0.085	0.86
2811a	HGG	44.87	0.021	0.044	0.038	0.63
2812a	HGG	62.73	0.041	0.087	0.171	0.68

Table 11 also lists the expanded prediction uncertainty, $U_{95,P}(H)$, which was computed as follows:

$$U_{95,P}(H) \ = \ Q_t(0.975, 18-1) \times \left(\tilde{\tau}^2 \ + \ u^2(\widetilde{H}) + (13+1)v^2\right)^{½}, \qquad \textbf{(7)}$$

where v denotes the geometric average of the standard uncertainties associated with the block-specific values of average hardness:

$$v = exp(\textstyle\sum_{j=1}^{18} \log u_c \, (\overline{H}_j) \ / \ 18). \qquad \textbf{(8)}$$

This evaluation of $U_{95,P}(H)$ is only a rough approximation, because it neglects correlations between block-specific values of average hardness and the estimates of both the consensus value and of τ.

The expanded uncertainty, $U_{95}(\widetilde{H})$, that qualifies a value of hardness, $\widetilde{H}$, assigned to an SRM expresses NIST's belief, with 95 % probability, that the true value of the average of the block-specific true values of average hardness lies in the interval $\widetilde{H} \pm U_{95}(\widetilde{H})$.

The prediction expanded uncertainty, $U_{95,P}(H)$, can be interpreted as follows: if a block is selected at random from an SRM, and its hardness is measured at some location sufficiently far apart from preexisting indentations on that block using NIST's test method and a Rockwell hardness machine comparable to the NIST PHSM used for the certification of the material, then the resulting hardness value is expected, with 95 % probability, to lie in the interval $\widetilde{H} \pm U_{95,P}(H)$.

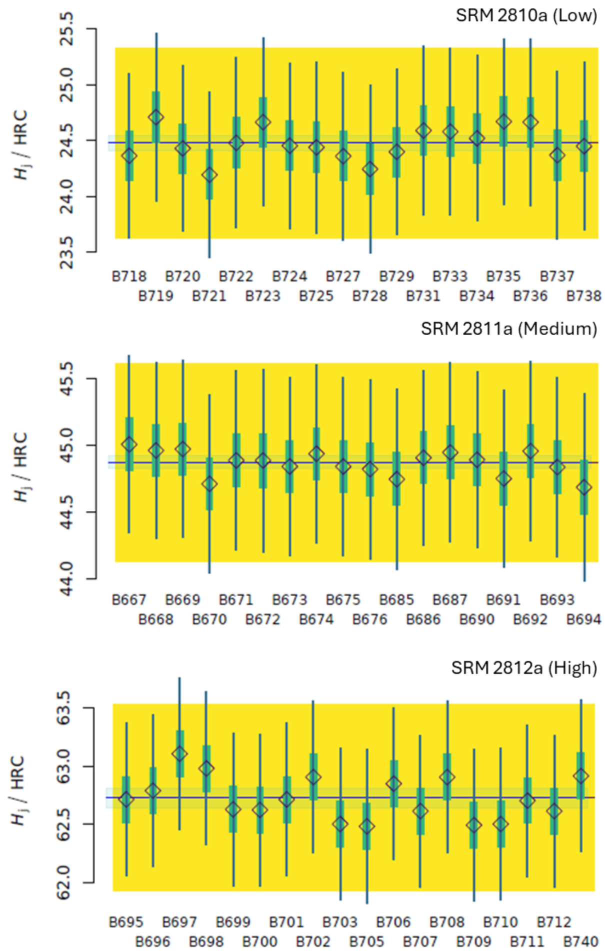


Figure 18. Block-specific estimates of the average hardness $\overline{H}_1, \ldots, \overline{H}_n$ of the n = 18 blocks in each SRM, are indicated by diamonds. The thick (green) vertical line segments represent $\overline{H}_j \pm U_{95}(\overline{H}_j)$, and the thin (dark blue) vertical line segments represent $\overline{H}_j + U_{95,P}(H_j)$ for j = 1, …, n. The thin horizontal line indicates the consensus value, $\widetilde{H}$, and the heights of the light blue and yellow bands represent confidence, $\widetilde{H} \pm U_{95}(\widetilde{H})$, and prediction, $\widetilde{H} \pm U_{95,P}(H)$, bands for the SRMs.

A. Worked Example for Comparing User Measurements with Certified Values

A.1. Scope

This appendix provides an informational worked example for comparing a user's Rockwell C hardness (HRC) measurement results with the certified value of a Rockwell C Standard Reference Material (SRM). The example illustrates how Welch's t test may be used to evaluate whether the observed difference between the user's average measured hardness value and the certified hardness value is statistically significant, considering the associated standard uncertainties and effective degrees of freedom.

This appendix is intended to illustrate the calculation and interpretation steps. It does not establish additional requirements for use of the Rockwell C SRM, nor does it prescribe an acceptance criterion.

A.2. Example Scenario and User Measurements

Suppose a user wishes to verify her hardness measurements using block B718 (**Table 8 of Section 9.3.1**). The user makes $m = 5$ indentations with her machine, all sufficiently far from pre-existing indentations on the SRM block surface, and employs the measurement procedure used routinely in her laboratory.

Let $H_{m,i}$ denote the ith Rockwell C hardness value measured by the user, where $i = 1, \dots, m$. The individual values, $H_{m,i}$, are listed in **Table A.1** and are used to calculate the user average hardness, $\overline{H}_m$, and its associated standard uncertainty, $u(\overline{H}_m)$.

Table A.1. Summary of measurement and certification values used in the example.

Symbol	Value	Description
m	5	Number of user indentations
$H_{m,i}$	25.22, 24.84, 25.11, 25.33, 24.05 (all in HRC)	Individual user hardness values, where $i = 1, \dots, m$
$\overline{H}_m$	24.91 HRC	Average hardness of user measurements
$u(\overline{H}_m)$	0.23 HRC	Standard uncertainty of the user average
ν_m	4	Degrees of freedom associated with the user's average hardness, equal to $m - 1$
$\overline{H}_C$	24.36 HRC	Certified hardness value
$u(\overline{H}_C)$	0.09 HRC	Standard uncertainty associated with the certified hardness value
ν_C	12	Degrees of freedom associated with the certified hardness value
ν_{EFF}	5.3	Effective degrees of freedom from the Welch-Satterthwaite formula [30, Eq. (2.44)]
p	0.07	Two-sided probability value, or p-value, from Welch's t test

The average hardness is calculated as:

$$\overline{H}_m = \frac{1}{m}\sum_{i=1}^{m} H_{m,i} = \frac{25.22 + 24.84 + 25.11 + 25.33 + 24.05}{5} = 24.91\ \mathrm{HRC}.$$

(A.1)

The associated standard uncertainty of the user average is the sample standard deviation, s_m, of the m determinations divided by the square root of the number of determinations:

$$u(\overline{H}_m) = \frac{s_m}{\sqrt{m}} = 0.23\ \mathrm{HRC}.$$

(A.2)

The degrees of freedom associated with the user average are:

$$\nu_m = m - 1 = 4.$$

(A.3)

A.3. Welch's *t* Test

The user performs Welch's t test [30, Eq. (2.38)] based on the standardized difference between the average of her five determinations and the block's certified hardness.

The difference between the user's average hardness and the certified value is:

$$\Delta = \overline{H}_m - \overline{H}_C = 0.55\ \mathrm{HRC}.$$

(A.4)

Using $u(\overline{H}_m) = 0.23\ \mathrm{HRC}$ and $u(\overline{H}_C) = 0.09\ \mathrm{HRC}$, the standard uncertainty associated with this difference is:

$$u(\Delta) = \sqrt{u^2(\overline{H}_m) + u^2(\overline{H}_C)} = 0.25\ \mathrm{HRC}.$$

(A.5)

The standardized difference is calculated as

$$t = \frac{|\overline{H}_m - \overline{H}_C|}{\sqrt{u^2(\overline{H}_m) + u^2(\overline{H}_C)}} = \frac{|\Delta|}{u(\Delta)} = 2.23.$$

(A.6)

A.4. Effective Degrees of Freedom

The reference probability distribution for this statistic is Student's t distribution with effective degrees of freedom determined using the Welch-Satterthwaite formula [30, Eq. (2.44)]. Substituting $u(\overline{H}_m) = 0.23$ HRC, $u(\overline{H}_C) = 0.09$ HRC, $\nu_m = 4$, and $\nu_C = 12$ gives:

$$\nu_{\text{EFF}} = \frac{\left[u^2(\overline{H}_m) + u^2(\overline{H}_C)\right]^2}{\frac{u^4(\overline{H}_m)}{\nu_m} + \frac{u^4(\overline{H}_C)}{\nu_C}} = 5.3.$$

(A.7)

A.5. Interpretation of Results

The *p*-value is determined from the standardized difference in Equation (A.6) and the effective degrees of freedom in Equation (A.7). Because the comparison asks whether the user's average hardness differs from the certified value in either direction, a two-sided Student's t test is used.

For this example, the *p*-value is the probability that a Student's t random variable with $\nu_{\text{EFF}} = 5.3$ degrees of freedom has an absolute value at least as large as the observed standardized difference, $|t| = 2.23$:

$$p = \Pr\left(\left|T_{\nu_{\text{EFF}}}\right| \geq |t|\right) = \Pr(|T_{5.3}| \geq 2.23).$$

(A.8)

Equivalently, using the cumulative distribution function of Student's t distribution,

$$p = 2[1 - F_t(|t|; \nu_{\text{EFF}})] = 2[1 - F_t(2.23; 5.3)] = 0.07.$$

(A.9)

Thus, the *p*-value of the test is approximately 0.07. In this context, a probability of about 7 % is not sufficiently small to provide strong evidence of a statistically significant difference between the user's hardness measurement result and the certified hardness value. For example, relative to a conventional 5 % significance level, $p = 0.07$ would not be considered statistically significant. However, the applicable decision criterion should be selected by the user or laboratory, as appropriate.

The user therefore concludes that her measurement of the block's average hardness is consistent with the certified value and does not differ significantly from it.

A.6. Note on Prediction Expanded Uncertainty

To reach the conclusion in Section A.5, the user did not use the prediction expanded uncertainty listed in **Table 8** of **Section 9.3.1**. For block B718, **Table 8** gives the certified block average hardness as $\overline{H}$ = 24.36 HRC and the 95 % expanded prediction uncertainty as $U_{95,P}(H) = 0.73\ HRC$.

The corresponding 95 % prediction interval for a single indentation result is obtained by applying the prediction expanded uncertainty to the certified block average hardness:

$$H = \overline{H} \pm U_{95,P}(H) = 24.36 \pm 0.73\, HRC = [23.63, 25.09]\, HRC.$$

(A.10)

Thus, if a single indentation is made on the SRM block surface using a machine and measurement method essentially similar to NIST's, the result is expected to lie between 23.63 HRC and 25.09 HRC with an approximately 95 % probability.

The prediction expanded uncertainty answers a different question from Welch's *t* test. The prediction interval applies to an individual indentation result, whereas Welch's *t* test compares the user's average hardness value with the certified hardness value, accounting for the uncertainties associated with both quantities.

Although three of the five user hardness determinations fall outside this 95 % prediction interval, this is not unexpected when the user's indentation results are assumed to be mutually independent. Under that assumption, the probability that at least one of five independent results falls outside a 95 % prediction interval is

$$1 - (0.95)^5 = 0.23.$$

(A.11)

Thus, there is about a 23 % probability that at least one of five independent results will fall outside the 95 % prediction interval. This note is included to clarify that the prediction expanded uncertainty and Welch's *t* test address different questions.

References

1. JCGM 2012 International Vocabulary of Metrology—Basic and General Concepts and Associated Terms *JCGM 200:2012* (2012). https://doi.org/10.59161/JCGM200-2012.
2. Bruce, S., Possolo, A., & Watters, R., Jr. (2021). Metrological traceability: Frequently asked questions and NIST policy, *NIST Technical Note 2156*. NIST. https://doi.org/10.6028/NIST.TN.2156.
3. ASTM E28.06. Standard Test Methods for Rockwell Hardness of Metallic Materials. Standard ASTM E18-25, *ASTM International*, West Conshohocken, PA, 2024.
4. ISO 6508-1:2023, Metallic Materials — Rockwell Hardness Test: part 1. Test Method.
5. ISO 6508-3:2023, Metallic Materials — Rockwell Hardness Test: part 3. Calibration of reference blocks.
6. BIPM Consultative Committee for Mass and Related Quantities (CCM) Working Group on Hardness (WGH). Rockwell Hardness HRC Scale Definition. International Bureau of Weights and Measures, (2021). https://www.bipm.org/documents/20126/63747533/HRC+Definition+%282021+WGH+approved%29.pdf/c3d67a3f-7017-6f12-5a01-bc4b25e80ffe.

7. G. Barbato, S. Desogus, and R. Levi, Design and performance of a dead-weight standard Rockwell hardness testing machine, *Journal of Testing and Evaluation* 6(4), 276-279 (1978). https://doi.org/10.1520/JTE10954J.
8. G. Barbato, A. Germak, and S. Desogus, The IMGC hardness standard machine. Description of the actual software and of proposed modifications, *IMGC Report* (1992). https://hdl.handle.net/11696/34121.
9. A. Germak, A. Liguori, and C. Origlia, Experience in the metrological characterization of primary hardness standard machines, *Proc. IMEKO TC5 HARDMEKO* (2007).
10. L. He and Z. Feng, The Rockwell hardness primary national machine according to the new international definition, *7th International Symposium Proc. SPIE* 8321 (2011). https://doi.org/10.1117/12.904619.
11. K. Herrmann, Investigation of a calibration device for the depth measurement in Rockwell hardness testing, Proc. *IMEKO World Congress* (2003). Download: PWC-2003-TC5-002.pdf
12. BIPM/COOMET, *COOMET.M.H-S1 Final Report* (Comparison of Rockwell hardness standard machines between PTB and KazInMetr) (2009). https://doi.org/10.1088/0026-1394/46/1A/07001.
13. S. R. Low, A. Germak, et al., Developing definitions of conventional hardness tests for use by national metrology institute, *Measurement: Sensors* (2021). https://doi.org/10.1016/j.measen.2021.100096.
14. A. Prato et al., Experimental determination of sensitivity coefficients of some influence parameters in Rockwell B, C, 15N, 30N and 45N, *Measurement: Sensors* (2024). https://doi.org/10.1016/j.measen.2024.101449.
15. J. D. Fidelus and M. Kożuchowski, GUM's Rockwell hardness standard machines after modernisation, *Acta IMEKO* 9(5), 240 - 246 (2020). http://dx.doi.org/10.21014/acta_imeko.v9i5.977.
16. C. Kuzu, E. Pelit, Ç. Doğan, et al, A new design and automation of Rockwell hardness standard machine, *Proc. IMEKO World Congress* (2015). Download: IMEKO-WC-2015-TC5-164.pdf.
17. K. Hattori, S. Takagi, H. Ishida and T. Usuda, Traceability system of Rockwell hardness c scale in Japan, *Proc. IMEKO TC5 HARDMEKO* (2004). Download: IMEKO-TC5-2004-011.pdf.
18. J. F. Song, S. Low, D Pitchure, A Germak, et al, Establishing a world-wide unified Rockwell hardness scale with metrological traceability, *Metrologia* 34(331), 331 - 342 (1997). https://doi.org/10.1088/0026-1394/34/4/6.
19. S. Low, Rockwell Hardness Measurement of Metallic Materials, *NIST Special Publication 960-5*, National Institute of Standards and Technology (2001).
20. ISO 6508-2:2023, Metallic Materials — Rockwell Hardness Test: part 2. Verification and calibration of testing machines and indenters.
21. J. F. Song, S. Low, and A. Zheng, Geometric measurement comparisons for Rockwell diamond indenters. *Proc. IMEKO World Congress* (2009). Download: http://www.imeko2009.it.pt/Papers/FP_29.pdf.

22. Analytical Methods Committee. Robust Statistics — How Not to Reject Outliers. Part 1. Basic Concepts. *The Analyst*, 114:1693–1697 (1989). https://doi.org/10.1039/AN9891401693.
23. A. Possolo, A. Koepke, D. Newton, and M. R. Winchester. Decision Tree for Key Comparisons. *Journal of Research of the National Institute of Standards and Technology*, 126:126007 (2021). https://doi.org/10.6028/jres.126.007.
24. S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3,4):591–611 (1965). https://doi.org/10.2307/2333709.
25. J. L. Hodges and E. L. Lehmann. Estimates of location based on rank tests. *The Annals of Mathematical Statistics*, 34(2):598–611, (1963). https://doi.org/10.1214/aoms/1177704172.
26. S. Stoudt, A. Pintar, and A. Possolo. Uncertainty evaluations from small datasets. *Metrologia*, 58:015014, (2021). https://doi.org/10.1088/1681-7575/abd372.
27. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2025. URL https://www.R-project.org/.
28. A. Possolo and O. Bodnar. Redefining standard measurement uncertainty. *Ukrainian Metrological Journal*, 2022(1):34–37, (2022). https://doi.org/10.24027/2306-7039.1.2022.258815.
29. W. Q. Meeker, G. J. Hahn, and L. A. Escobar. *Statistical Intervals: A Guide for Practitioners and Researchers*. John Wiley&Sons, Hoboken, NJ, 2nd edition, 2017. ISBN 978-0-471-68717-7.
30. R. G. Miller. *Beyond ANOVA, Basics of Applied Statistics*. Chapman & Hall/CRC, Boca Raton, FL, 1997. ISBN 978-0-412-07011-2.