Temperature System Calibration Process Improvement and Risk Mitigation

Frank G. Ramírez Irizarry
Master in Manufacturing Engineering
José A. Morales Ph.D.
Industrial and Systems Engineering
Polytechnic University of Puerto Rico

Abstract *— The project's main objective was to update the software that performed the data collection for the Systems Calibration process. It held the high*est *priority in the facility's Risk Registry due to the obsolete computers. The software had consistent communication issues with most equipment bottlenecks to the weekly throughput of calibrations. We aimed to fix the compatibility issue, thus removing the process from the Risk* Registry, *and fixing the communication issues. By fixing the communication issues, we expected the process means to increase. To validate our hypothesis, we performed a Multiple Regression Analysis to identify which equipment being calibrated in this process significantly impacts the process mean. Finally, perform a Hypothesis Test of the Mean for each equipment being calibrated and the population mean using the data collected before the changes as baselines. Courses involved in this study were: Introduction to Instrumentation, Industrial System Automation, Lean and Six Sigma.*

Key Terms *— Hypothesis Test of Mean, Lean Six Sigma, Metrology, Process Improvement*

INTRODUCTION

The project takes place in the Measurement Technology Manufacturing and Service. This industry is focused on manufacturing and service instruments that perform physical measurements such as Temperature, Pressure, Mass, Volume, Current, and Resistance, among others. These devices and equipment keep all sorts of production, research, and services afloat. From toy manufacturing to life-saving medicine development, and modern pharmaceutical to food manufacturing measurement technology is necessary for modern to have such rigorous quality standards.

The objective is to perform a Process Improvement activity in the System Calibration Process for temperature instruments. The issue is numerous communication errors with the software when performing calibrations through it with specific readout models. The software communicates inconsistently with Readout Model A and does not allow the process to begin; the reference readout has communication issues when the software tells it to collect measurement data, and Model B has problems performing multiple probe calibrations simultaneously. These are some of the most impacting issues that are affecting this process.

This process is also found in the Risk Registry List due to the growing concern that the software was developed for 16-bit processors; it runs on 32-bit computer but not on 64-bit. These 32-bit computers are considered obsolete; all new computers are running 64-bit processors. It is urgently necessary to update the software for compatibility with newer computers since the previously mentioned computers are limited and no longer in production.

The System Calibration Process uses software that was developed specifically for it. This software manages the instruments and equipment used in the calibration process, but it is operated by a calibration technician who oversees it. The process was developed to calibrate temperature instruments containing a temperature-sensing probe and a readout. The calibration technician is responsible for selecting the unit, setting up the test, and operating the software to perform the necessary steps to complete the calibration. The process can be divided into two general procedures: the Fixed-Point Cell method, which is used for Platinum Resistance Thermometers, and the Comparison method, which is performed with Thermistor Probes, the software identifies the method.

This process calibrates a range of different readouts and probe combinations. Most temperature

readouts can use multiple different probe types. For example, readout A can use Standard Platinum Resistance Thermometer (SPRT), Platinum Resistance Thermometer (PRT), and Thermistor, but readout B can use PRTs, Thermistors, and Thermocouples.

The output studied is weekly calibration performed; multiple variables impact the number of calibrations that can be done. There are at least three readouts, and each readout can use three different measurement probes. For our study, we are not accounting for probes since the readouts perform the communication causing the issues. This process output is affected by the time a calibration takes. If an instrument fails calibration and multiple attempts are made to adjust, it will delay the calibration process output. An instrument that has connection issues will also cause a delay in the process of calibration. Another variable that indirectly impacts the process is the fixed-point method calibration since this process only allows a single instrument to be processed simultaneously.

Our objective is to collect data on the instruments calibrated weekly and analyze the meaning of the different models to justify resource allocation to perform improvements. The focus is to find enough data to prove that the communications are the root cause of the reduced calibrations performed. Once the data has been collected, we will be working on developing what is known as a software patch or software update to correct the issues found. This update will also be developed in a compatible version for the 64-bit computers to remove the process from the Risk Registry List.

The expected improvements from this project are to increase the throughput of the calibration process. By fixing these issues, the process will also be able to consistently complete calibrations on a single run without needing to perform multiple attempts. We expect to fix important software issues that require a complete station reboot, which significantly impacts the process's throughput. We also hope the process mean increases and decreases the throughput variance.

The importance of this project is significant as it will indirectly also impact other industries. Since this is a Service Process, it is indirectly attached to the customers. Multiple industries use temperature probes and readouts for their processes, such as Medical Devices, pharmaceuticals, Hardware Development, Food Manufacturing, and Research Laboratories. If issues in a process that service these industries delay their time to receive their equipment, it can cause these customers to search for other calibration service laboratories.

BACKGROUND

The following topics are important to the understanding of the process, project, and impact that it has in society and other industries.

Metrology

Metrology is the science of measurement; scientists and engineers in this field are called Metrologists. A measurement numerically represents an object's characteristic physical form or action. It allows us to understand and see matter change when it is impossible for the naked eye to witness. Measurements and other terms such as precision, tolerance, accuracy, uncertainty, and maximum permissible error are used closely to understand the behavior and characteristics of measurement and can be found in The International Vocabulary of Metrology – Basic and general concepts and associated terms (VIM) [1] by the International Bureau of Weights and Measures (BIPM). Figure 1 [2] shows an example of some of the terms. You can have a very accurate instrument that isn't precise; in another instance, you could have a very precise instrument that isn't accurate. Ideally, you would want an instrument that can achieve both.

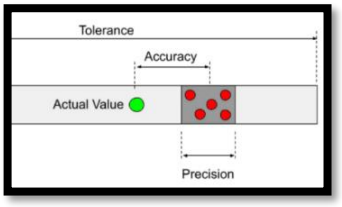

Figure 1
Comparison of Measurement Terms

Measurement Traceability

Measurement traceability is an unbroken chain of calibrations whose results can be traced back to the International System of Units through reference standards that each contribute a measurement uncertainty to the process, this calibration chain is shown below in Figure 2 [3].

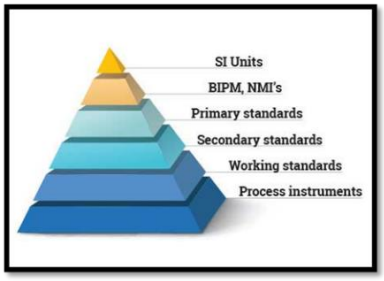

Figure 2
Pyramid of Traceability

Temperature Instrument

Temperature instruments are equipment that takes a physical value measured in degrees Fahrenheit, degrees Celsius, or degrees Kelvin and allows the end user to read these measured values. These instruments are often divided into two pieces of equipment: the probe and the readout. The temperature probe is the sensing portion of this instrument; it is the part exposed to the heating source the user wants to measure. The probe transduces the measured value as a change in its physical passive element. These passive electrical elements that change their behavior when exposed to the heating elements are constructed of different materials and components, such as thermistors and Platinum Resistance Thermometer (PRT)—both change resistance when exposed to temperature changes. As for the readout, this portion of the instrument displays the measurements. It is an end-user-focused device that takes the measured value from the probe and conditions these values to be shown in terms of temperature in a display.

Temperature Calibration

Calibration compares an instrument of an unknown value with a Reference Standard tested to have a known value. This process usually needs to be clarified with equipment adjustment performed after a calibration. An instrument calibration will not always merit an adjustment, but all adjustments are based on a calibration test. Calibration Technicians or Metrologists perform them in a controlled laboratory environment and have the necessary equipment to perform the task. They use what is known as a Reference Standard; this is another instrument that is far more precise than the instrument that will be tested and calibrated. These reference standards are calibrated in the most precise calibration laboratories in the world, and they are traceable to the International System of Units, the official system of measurement in the world. Various test points are established based on the instrument's measurement range in calibrating a user's instrument.

Usually, three test points are the minimum requirement: a low-test point in the measurement range, a high-test point in the measurement range, and one in between. We can compare the instrument's performance with these test points and determine its compliance. In temperature metrology, the reference standards usually are a Standard Platinum Resistance Thermometer (SPRT) probe paired with a high-precision readout and a Fixed-Point Cell or a Calibration Bath. The SPRTs are probes constructed with metal or glass sheaths covering platinum wire shown in Figure 3 below [4]. When its temperature changes, so does the resistivity of the platinum wire. The readout calculates this change, displaying the temperature measurement related to that resistance value.

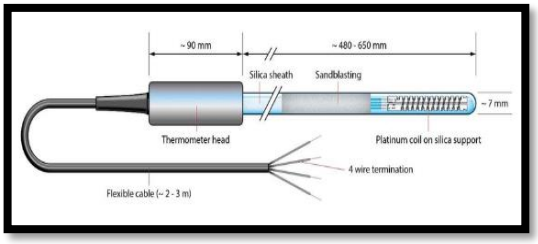

Figure 3
Construction of an SPRT

They are often used with instruments called Fixed-point cells at the highest calibration level. They are "well-like" glass instruments that enclose

an element substance and contain an aperture in the middle where the probe is introduced. These substances like Aluminum, Zinc, and Tin, among others, have a specific temperature point where their physical substance is altered. These cells are found inside furnaces designed to exert them to the necessary heat required to achieve their fixed point; in the case of the previously mentioned, the substance inside the cell changes from solid to liquid.

Other cells achieved their fixed point under other conditions, such as water and argon, where the temperature and their triple point are fixed. The triple point of an element is described as a value where the element contains its three states simultaneously: liquid, solid, and gaseous. Table 1 [5] below shows the column substance used, while the state column shows which form it will be in when in its fixed point. The temperature column shows the value at which that fixed point is reached in terms of Kelvins or ˚C.

Table 1
Substances and their Fix Point

Table 1. Defining fixed points of the ITS-90

Number	Temperature T_{90}/K	Temperature t_{90}/°C	Substance[a]	State[b]	$W_r(T_{90})$
1	3 to 5	−270,15 to −268,15	He	V	
2	13,8033	−259,3467	e-H_2	T	0,001 190 07
3	≈17	≈−256,15	e-H_2 (or He)	V (or G)	
4	≈20,3	≈−252,85	e-H_2 (or He)	V (or G)	
5	24,5561	−248,5939	Ne	T	0,008 449 74
6	54,3584	−218,7916	O_2	T	0,091 718 04
7	83,8058	−189,3442	Ar	T	0,215 859 75
8	234,3156	−38,8344	Hg	T	0,844 142 11
9	273,16	0,01	H_2O	T	1,000 000 00
10	302,9146	29,7646	Ga	M	1,118 138 89
11	429,7485	156,5985	In	F	1,609 801 85
12	505,078	231,928	Sn	F	1,892 797 68
13	692,677	419,527	Zn	F	2,568 917 30
14	933,473	660,323	Al	F	3,376 008 60
15	1234,93	961,78	Ag	F	4,286 420 53
16	1337,33	1064,18	Au	F	
17	1357,77	1084,62	Cu	F	

Temperature calibration is often divided into two process methods: Comparison and Fixed-Point. In a Comparison method, you will have a bath [6] or a dry block [7] where the temperature can be controlled to a desired set point for calibration, and multiple probes can be inserted. These baths are far less precise than a fixed point; that is when the calibration technicians use a Reference Standard. The method is called comparison because the calibration technician places the reference standard and the instrument under test in the heat source, whether an oil bath or dry block. The temperature is taken with the reference standard, and the temperature measurement taken with the instrument under test will be compared to that of the reference standard, based on that comparison and the precision specification of the instrument under test, it will determine if it is compliant.

Oil baths contain a temperature controller, a heating element, a stirrer, and oil. The temperature controller monitors the set point and the actual temperature, actioning the heating element to increase its value or decrease; the stirrer is used to mix the oil, ensuring that all surface oil reaches the same temperature. Generally, dry blocks are cylindrical metal inserts that contain orifices for the probes to be inserted; they also contain a temperature controller and a heating element. In this case, a stirrer and oil are unnecessary as the metal insert is exerted to heat and cover the probes that will be tested. An example of comparison calibration using a dry block is shown here in Figure 4 [8].

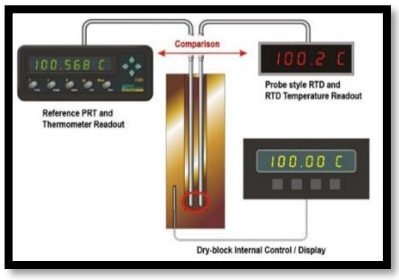

Figure 4
Comparison Calibration

In a Fixed-Point method, a substance cell is placed in a Furnace that will control its required temperature, the construction of this furnace is shown in Figure 5 [9]. The reference standard is used to measure and verify the temperature of the Fixed-Point Cell to ensure that its required temperature is met and stable. The calibration technician keeps track of the cell measurement, periodically inserting the reference standard probe before inserting the instrument under test. The instrument under test is inserted, and the recorded temperature is taken. Based on the temperature taken of the Fixed-point Cell with the reference standard probe, the temperature measured and the precision specification of the instrument under test are compared and verified if it complies.

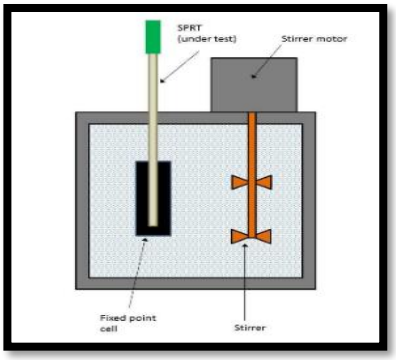

Figure 5
Fixed Cell Furnace

The difference between these two methods is that in a comparison test, the accuracy of the process is significantly lower. Therefore, it is a process for lower accurate probes, and because the number of probes that can be inserted at once for a calibration test point is higher, this is a volume-biased process. On the other hand, the fixed-point cell method is reserved for high-end accurate probes that require the most precise calibration process. Because of their singular aperture in the cell's well, only one probe can be calibrated at a test point at a time. Figure 6 below [10] is an example of the construction of a fixed-point cell. Its limitations are the single well, the delicate construction comprised mainly of glass, and the development complexity. They require very high care, and a handful are manufactured at a given time. If this equipment breaks, it can set back a calibration process for high downtime.

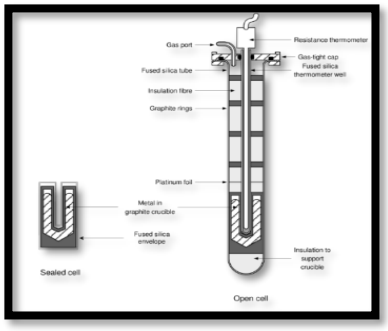

Figure 6
Fixed Point Cell Construction

Lower precision probes are used in environments where monitoring temperature is important but not crucial. For example, the polymer melts at a specific temperature in the molding process of a polymer used for phone cases. You will want to monitor the temperature of the mold to ensure that it has reached its set point, but it is not crucial for the plastic to not strive for its temperature, let's say 0.01˚C off. On the other hand, higher precision probes are used in strict control environments where every decimal point is crucial to process output. An example of industries that require these types of instruments are Food Manufacturing and Pharmaceutical industries. These industries need their measurements to be precise and true to the process. They are working with consumer goods that will be ingested, and even the slightest temperature that is off specification for their process can cause problems for a batch and its consumers. Perhaps the bacterium in a product is not completely sterilized if the process temperature is off by 0.1˚C. They must also provide proper information and documentation of their process to federal agencies that oversee their operations, such as the Food and Drug Administration (FDA). Such as monitored data of the process, process inspection, and calibration certificate of an accredited laboratory, among others.

Calibration Adjustment

Generally, when equipment is calibrated, it is not necessarily adjusted. Instrument adjustment is the action of ensuring that measurements made by the equipment lie within the accuracy of it. If a calibration process is made and the instrument is found to be outside the established accuracy specifications, then an adjustment is performed.

Lean Six Sigma

Lean Six Sigma combines two process improvement methodologies, Lean and Six Sigma. Lean methods focus on process improvement by removing process waste as much as possible or reducing it. There are eight types; an in-depth description of each waste can be seen in Figure 8 below [11]. Our process is affected by waste related to waiting and defects as it relates to wait times when a station faces an issue that requires a machine reboot.

Figure 8
Types of Waste

The Six Sigma methodology focuses on reducing process variation and increasing its control. It aims to achieve it using statistical analysis of a process evaluating the variance of the variables that impact its output. A commonly used method in Six Sigma is DMAIC (Define, Measure, Analyze, Improve, and Control) its road map can be seen in Figure 9 below [12].

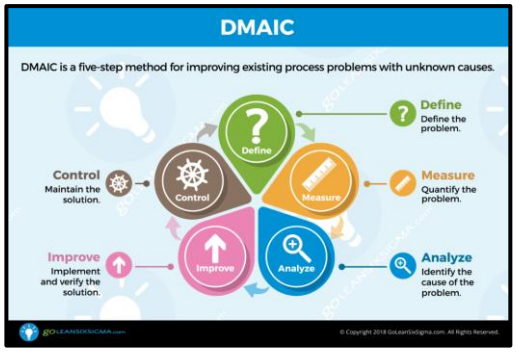

Figure 9
DMAIC Process

With Six Sigma, we are working with data bias decision-making to strive to reduce process variance using the DMAIC method. By implementing this methodology in our process, we expect to minimize the variance of software communication issues impacting the calibration. By doing so, we can efficiently map calibration intervals for the instruments and estimate a shorter and more consistent calibration elapse.

While Lean focuses on eliminating or reducing waste, Six Sigma focuses on the variation of process variables affecting its output. The combination of Lean-Six Sigma, a powerful problem-solving tool that helps its user focus their attention on the root cause. It helps improve the workflow of a process and its quality. The following Figure 10 is a representation of the independent methodologies and their objectives, including their combination [13].

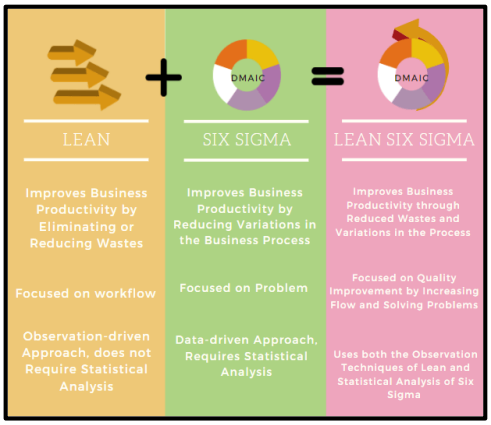

Figure 10
Lean-Six Sigma Methodology

Our objective is to reduce or eliminate the downtime waste of the stations by fixing the software communication problems that inhibit the process workflow from proceeding with its course. While making these changes, we will also expect to eliminate the obsolescent waste created by old machines running on 32-bit systems and convert the software to a 64-bit compatible program. In the efforts to improve and maintain the operations of the process, we are considering implementing control charts to monitor software issues to alert potential risks that the process might be against in the future when other computer and equipment drivers are updated.

METHODOLOGY

Twelve samples were taken, each corresponding to a week of calibration for the System Calibration Process. During these twelve weeks, we divided and grouped the calibrations performed by the different readouts. Every readout in a population refers to a base model number in which the variations of those models were not accounted as different equipment. We recovered a total mean of the population of 6.5 units calibrated per week, with a Standard deviation of 6.5. Readout A had a weekly mean of 1.9 Units with a Standard deviation of 2.7, Readout B had the highest mean with 12.6 and a standard deviation of 7.2, and Readout C had a mean of 5.1 and a standard

deviation of 2.8. Below is the descriptive statistic in Table 2 of the data collected using Minitab Statistical Software.

Table 2

Descriptive Statistics of the Data Using Minitab

Descriptive Statistics: Readout A, Readout B, Readout C

Statistics

Variable	N	N*	Mean	SE Mean	StDev	CoefVar	Sum	Minimum	Q1	Median	Q3
Readout A	12	0	1.917	0.793	2.746	143.24	23.000	0.000	0.000	1.000	3.000
Readout B	12	0	12.58	2.07	7.18	57.05	151.00	1.00	7.50	12.00	19.50
Readout C	12	0	5.083	0.821	2.843	55.93	61.000	1.000	3.250	4.500	7.000

Variable	Maximum
Readout A	9.000
Readout B	24.00
Readout C	10.000

We performed a multiple regression analysis using the weekly Mean and two of the readouts to have enough residuals to collect P-Values [14]. We used an alpha of 0.05 on all our multiple regression analyses. Pairing two readout data while leaving one for residuals, we made three multiple regression analyses to cover all bases to identify which is of significant impact on our weekly calibration mean. Our first multiple regression analysis covers Readout A and Readout C. Table 3 below shows that with this interaction, we obtain a P-Value for Readout A of 0.360 and a P-Value for Readout C of 0.074. They do not affect our weekly calibration means by themselves.

Table 3

Multiple Regression Analysis of Readout A and C

Regression Analysis: Weekly Mean versus Readout A, Readout C

Regression Equation

Weekly Mean = 3.22 + 0.271 Readout A + 0.548 Readout C

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	3.22	1.68	1.92	0.087	
Readout A	0.271	0.280	0.97	0.360	1.00
Readout C	0.548	0.271	2.02	0.074	1.00

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
2.54737	34.70%	20.19%	0.00%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	31.033	15.517	2.39	0.147
Readout A	1	6.043	6.043	0.93	0.360
Readout C	1	26.590	26.590	4.10	0.074
Error	9	58.402	6.489		
Lack-of-Fit	7	50.124	7.161	1.73	0.414
Pure Error	2	8.278	4.139		
Total	11	89.435			

At our next regression analysis, we looked over Readouts A and B. Table 4 below shows that the P-values recovered for this interaction were Readout A with 0.019 and Readout B with 0.000. With the data recovered, we can assume that they impact our weekly calibration mean.

Table 4

Multiple Regression Analysis of Readout A and B

Regression Analysis: Weekly Mean versus Readout A, Readout B

Regression Equation

Weekly Mean = 1.299 + 0.317 Readout A + 0.3671 Readout B

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	1.299	0.662	1.96	0.081	
Readout A	0.317	0.111	2.85	0.019	1.01
Readout B	0.3671	0.0426	8.62	0.000	1.01

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
1.01051	89.72%	87.44%	81.49%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	80.245	40.123	39.29	0.000
Readout A	1	8.290	8.290	8.12	0.019
Readout B	1	75.802	75.802	74.23	0.000
Error	9	9.190	1.021		
Total	11	89.435			

In the following regression analysis, we looked over Readouts B and C. Table 5 below shows the P-values recovered for this interaction were Readout B with a value of 0.000 and Readout C with 0.018. With the data recovered, we can assume that they impact our weekly calibration mean.

Table 5

Multiple Regression Analysis of Readout B and C

Regression Analysis: Weekly Mean versus Readout B, Readout C

Regression Equation

Weekly Mean = 0.842 + 0.3235 Readout B + 0.318 Readout C

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.842	0.737	1.14	0.282	
Readout B	0.3235	0.0438	7.39	0.000	1.07
Readout C	0.318	0.111	2.87	0.018	1.07

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
1.00656	89.80%	87.54%	84.32%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	80.317	40.158	39.64	0.000
Readout B	1	55.326	55.326	54.61	0.000
Readout C	1	8.361	8.361	8.25	0.018
Error	9	9.119	1.013		
Total	11	89.435			

The consensus is to improve the weekly calibration, which means all process changes must positively impact Readouts B. Let's consider the first regression analysis in Table 3. Changes that only benefit Readouts A and C will not be significant to the weekly mean if they do not support an indirect benefit to the number of calibrations performed to Readouts B. By improving the communication issues that are bottlenecking Readouts A the freed time should be used to boost calibrations for Readouts B.

After successfully updating the software to be compatible with 64-bit computing operating systems and fixing most of the communication issues the software had with the readouts, we went back to data collection. Once again, we focused on weekly calibrations performed and aimed for another twelve-week data collection interval. After collecting as much data as possible, we went back to evaluating the data. We performed another Descriptive Statistics run using Minitab, shown in Table 6 below. For Readout A, the mean recovered was 0.25 with a standard deviation of 0.622, Readout B had a mean of 5.00 with a standard deviation of 6.58, and Readout C had a mean of 2.42 with a standard deviation of 3.75.

Table 6
Descriptive Statistics of the Data Collected After Improvements using Minitab

Descriptive Statistics: Readout A After, Readout B After, Readout C After

Statistics

Variable	N	N*	Mean	SE Mean	StDev	CoefVar	Sum	Minimum	Q1	Median	Q3
Readout A After	12	0	0.250	0.179	0.622	248.63	3.000	0.000	0.000	0.000	0.000
Readout B After	12	0	5.00	1.90	6.58	131.56	60.00	0.00	0.00	0.00	12.25
Readout C After	12	0	2.42	1.08	3.75	155.29	29.00	0.00	0.00	0.00	5.75

Variable	Maximum
Readout A After	2.000
Readout B After	17.00
Readout C After	10.00

RESULTS AND ANALYSIS

Once the changes were made and the communication was consistent, we collected weekly calibrations performed. Once we had enough weekly calibration data, we performed our statistical analysis. We proceeded to perform the Mean Hypothesis test for Two samples using Student T. We used Student T because for the separate hypothesis mean testing, the readouts samples are twelve, below a thirty-sample size that would be best used in a Z-Test. To determine if there was an improvement in the means of each readout, we used Minitab to perform the two-sample t-tests with an alpha of 0.05. We established that μ_1 = Readout After Improvement mean and that μ_2 = Readout Before Improvement Mean. Since we wanted to know if there was an increase in the calibration mean, the difference of $\mu_1 - \mu_2$ should be Greater than zero. Therefore, our H_o or Null Hypothesis was Difference = 0, and our H_a or Alternate hypothesis would be Difference > 0. Table 7 below shows the test performed for Readout A, with a P-Value of 0.969. We fail to reject the null hypothesis; there is insufficient evidence to conclude that our Readout A calibration mean has increased.

Table 7
Mean Hypothesis Test for Readout A, Before and After Improvements

Two-Sample T-Test and CI: Readout A After, Readout A Before

Method

μ_1: population mean of Readout A After
μ_2: population mean of Readout A Before
Difference: $\mu_1 - \mu_2$

Equal variances are not assumed for this analysis.

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
Readout A After	12	0.250	0.622	0.18
Readout A Before	12	1.92	2.75	0.79

Estimation for Difference

Difference	95% Lower Bound for Difference
-1.667	-3.115

Test

Null hypothesis H_0: $\mu_1 - \mu_2 = 0$
Alternative hypothesis H_1: $\mu_1 - \mu_2 > 0$

T-Value	DF	P-Value
-2.05	12	0.969

Table 8 below shows the test performed for Readout B, with a P-value of 0.993. We fail to reject the null hypothesis; there is insufficient evidence to conclude that our Readout B calibration mean has increased.

Table 8
Mean Hypothesis Test for Readout B, Before and After Improvements

Two-Sample T-Test and CI: Readout B After, Readout B Before

Method

μ_1: population mean of Readout B After
μ_2: population mean of Readout B Before
Difference: $\mu_1 - \mu_2$

Equal variances are not assumed for this analysis.

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
Readout B After	12	5.00	6.58	1.9
Readout B Before	12	12.58	7.18	2.1

Estimation for Difference

Difference	95% Lower Bound for Difference
-7.58	-12.42

Test

Null hypothesis H_0: $\mu_1 - \mu_2 = 0$
Alternative hypothesis H_1: $\mu_1 - \mu_2 > 0$

T-Value	DF	P-Value
-2.70	21	0.993

Table 9 below shows the test performed for Readout C, with a P-value of 0.968. We fail to reject the null hypothesis; there is insufficient evidence to conclude that our Readout C calibration mean has increased.

Table 9
Mean Hypothesis Test for Readout C, Before and After Improvements

Two-Sample T-Test and CI: Readout C After, Readout C Before

Method

μ_1: population mean of Readout C After
μ_2: population mean of Readout C Before
Difference: $\mu_1 - \mu_2$

Equal variances are not assumed for this analysis.

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
Readout C After	12	2.42	3.75	1.1
Readout C Before	12	5.08	2.84	0.82

Estimation for Difference

Difference	95% Lower Bound for Difference
-2.67	-5.01

Test

Null hypothesis H_0: $\mu_1 - \mu_2 = 0$
Alternative hypothesis H_1: $\mu_1 - \mu_2 > 0$

T-Value	DF	P-Value
-1.96	20	0.968

To determine if there was an improvement in the System Calibration process mean, we used Minitab to perform the one-sample Z-Test with an alpha of 0.05. Since we know the population's standard deviation and a sample size greater than thirty, we proceeded with the Z-Test. We established that $\mu = 6.5$ as our H_o or Null Hypothesis since that was the process's mean before the improvement. Since we wanted to know if there was an increase in the calibration mean, our H_a or Alternate hypothesis would be $\mu > 6.5$. Table 10 below shows the test performed for the calibration process mean, with a P-value of 1.000. We fail to reject the null hypothesis; there is insufficient evidence to conclude that our calibration process mean has increased.

Table 10
Mean Hypothesis Test for the System Calibration Process After Improvements

One-Sample Z

Descriptive Statistics

N	Mean	SE Mean	95% Lower Bound for μ
36	2.600	0.783	1.312

μ: population mean of Sample
Known standard deviation = 4.7

Test

Null hypothesis H_0: $\mu = 6.5$
Alternative hypothesis H_1: $\mu > 6.5$

Z-Value	P-Value
-4.98	1.000

DISCUSSION AND CONCLUSION

After performing Hypothesis Tests of Mean for each readout independently and the process mean, we can state that there was insufficient evidence to conclude that any of the means had increased in response to the software update and communication fixes. Due to many unexpected issues during the data collection after the changes, the data collected certainly does not represent a regular process operation. Issues that directly affected this process were when Station 1, using the old 32-bit software, stopped working, and the installation of the 64-bit version used in this project development had to be implemented. The throughput was still reduced due to the allocation time for a new 64-bit machine and the setup time to have Station 1 continue servicing. Another issue was that the calibration baths used in this process experienced higher than allowed measurement uncertainty, resulting in a processing halt for investigation. Issues that indirectly affected the process were personnel limitations, such as other laboratories requiring calibration technicians to assist their process and alleviate the backlog, which resulted in a lower throughput of the calibrations performed.

Summary of contributions

This software project held one of our facility's top scores in the Risk Registry. The uncertainty in terms of time needed to know if the program would be able to be made compatible with current 64-bit computers was vast. If the software code available to us could not be compatible, reverse engineering would be contemplated and would take a long time to develop. In terms of Risk Mitigation, the project was called a success. When Station 1 went down, we had a running program that could be implemented quickly and significantly helped with its downtime.

Future Recommendations

We recommend performing the mean test once again with a greater sample size and after the operations in the laboratory have normalized. It is also recommended to periodically verify software

compatibility with future operating system updates to reduce the risk of running software before its compatible operating system is obsolete. Any software over two–three years must be evaluated to ensure that software frameworks will still be supported on future operating systems.

REFERENCES

[1] BIPM (2012). *International vocabulary of metrology – Basic and general concepts and associated terms (VIM)* (3rd ed.). JCGM. Available: https://www.bipm.org/documents/20126/2071204/JCGM_200_2012.pdf/f0e1ad45-d337-bbeb-53a6-15fe649d0ff1?version=1.16&t=1659082802818&download=true

[2] T. X. (2021, October 2). What is Metrology and How is it Used in Manufacturing? Xometry. [Online] https://www.xometry.com/resources/blog/metrology/

[3] F. (n.d.). Learn About Calibration Basics. Fluke Calibration. [Online] https://us.flukecal.com/literature/about-calibration

[4] (2021). Guide to the Realization of the ITS-90 Platinum Resistance Thermometry (p. 16). BIPM. [Online] https://www.bipm.org/en/committees/cc/cct/guides-to-thermometry

[5] Perston-Thomas, H. (1989, October 24). The International Temperature Scale of 1990 (ITS-90). Metrologia, 27(3-10), 5.

[6] A. (2022). Standard Guide for the Preparation and Evaluation of Liquid Baths Used for Temperature Calibration by Comparison. ASTM.

[7] A. (2019). Standard Guide for Use and Testing of Dry-Block Temperature Calibrators. ASTM.

[8] Potter, S. (2018, July 24). How to do a temperature sensor comparison calibration. Fluke Calibration. https://us.flukecal.com/blog/how-do-temperature-sensor-comparison-calibration

[9] Joung, W. (2015, December 1). Guide to the Realization of the ITS-90: Metal Fixed Points for Contact Thermometry. Research Gate. Available: https://www.researchgate.net/figure/Typical-stirred-liquid-bath_fig5_325075540

[10] Machin, G. (2017, December 1). Guide on secondary thermometry: Specialised fixed points above 0°C. Research Gate. Available: https://www.researchgate.net/figure/Simplified-diagram-showing-the-construction-of-metal-fixed-point-cells-not-to-scale_fig4_321885305

[11] Skhmot, N. (2017, August 5). The 8 Wastes of Lean. The Lean Way. [Online] https://theleanway.net/The-8-Wastes-of-Lean

[12] Lean Six Sigma: Step by Step (DMAIC Infographic). (n.d.) [Online] Goleansixsigma.com. https://goleansixsigma.com/lean-six-sigma-step-by-step/

[13] K. (2021, July 31). Lean vs Six Sigma: Understanding the Lean Six Sigma Methodology. UNICHRONE. https://unichrone.com/blog/quality-management/lean-vs-six-sigma/

[14] Beers, B. (2023, March 28). *P-Value: What It Is, How to Calculate It, and Why It Matters*. Investopedia. https://www.investopedia.com/terms/p/p-value.asp