This document was prepared in conjunction with work accomplished under Contract No. DE-AC09-96SR18500 with the U. S. Department of Energy.

REPRODUCIBILITY: A MAJOR SOURCE OF UNCERTAINTY IN WEIGHING[1]

By

John P. Clark & A. Harper Shull
Westinghouse Savannah River Company
Aiken, S.C. 29808 USA

Phones: 803-725-3654 & 803-952-3577
Email: & johnp.clark@srs.gov **&** harper.shull@srs.gov

ABSTRACT

Balance confirmation tests usually include eccentricity, linearity, and repeatability. The latter is a measure of the random variability of the instrument. Further, repeatability is usually based on 10 consecutive measurements of a mass standard in a short period of time. The test results are used to verify the balance performs within manufacturer's specifications. These test results are often used in calculating an estimate of the uncertainty in measurements made with the balance. Uncertainty estimates, used to quantify the quality of measurements, must accurately estimate the magnitude of these errors. Does the repeatability test provide a realistic estimate of the random variation in the weighing process? The authors had concerns that it may not. The reproducibility estimate of the balance needs to be determined for the variables that will change between measurements of the same material while the other parameters are held constant by procedures and training.

Five experiments were conducted over a two year period to determine the magnitude of random and systematic errors associated with weighing known mass standards. Repeated tests showed significant day to day variation (*reproducibility*) in the averages of the balance repeatability measurements. The first set of repeatability test averages varied by more than 10 scale divisions between some tests. The standard deviation of approximately 200 measurements exceeded the repeatability specification listed by the manufacturer's repeatability specification by a factor of 3 in the first phase of the experiments. The results of the first experiment led to subsequent tests. The results indicate a significant portion of the variation is due to the "between" days variation. Also, the larger the quantity of material weighed on the balance, the greater the day to day portion of the total variation. There is a strong correlation between humidity and the variation of the daily average values. Therefore, it is important to determine the magnitude of day to day variation and include it the calculation of the uncertainty estimate for quantity of material being weighed. The evaluation of QA program data for several balances and different check standards revealed the standard deviation of all the measurements taken over several weeks, far exceeded the "repeatability" specifications for different electronic balances. Estimates of uncertainty should be based on "reproducibility" estimates, using check standards approximating the magnitude of the material weighed rather than the specifications for the balance or the standard deviation of a repeatability test.

Introduction

A new 5 place precision balance was purchased for pipette calibration experiments. The manufacturer provided a calibration certificate that included test results, which indicated the sensitivity adjustment, eccentricity, linearity, and repeatability specifications, were met. The certificate did not include an uncertainty estimate. Repeatability, eccentricity and linearity tests were conducted to verify the balance performed within manufacturer's specifications at the user's location. The results verified the balance performed to manufacturer's specifications. However, when the repeatability test was conducted a second time with the same 100 g mass standard in making 10 measurements, it had a significantly different average value. This caused concern. Since the test was repeated under different environmental conditions the combination of several sets of repeatability test results provided data for estimating the "reproducibility" of measurements. It is useful to review the definitions of these two terms as given in the International Vocabulary of Basic and General Terms in Metrology Vocabulary[2].

> **3.6 repeatability (of results of measurements)** closeness of the agreement between the results of successive measurement of the same measurand carried out under the same conditions of measurement.
>
> **NOTES**
> 1. These conditions are called repeatability conditions.
> 2. Repeatability conditions include: the same measurement procedure, the same observer, the same measuring equipment used under the same conditions, the same location and repetition over a short period of time.
> 3. Repeatability may be expressed quantitatively in terms of the dispersion characteristics of the indications as a short-term random standard deviation.
>
> **3.7 reproducibility (of results of measurement)** closeness of the agreement between the results of measurement of the same measurand carried out under changed conditions of measurement.
>
> **NOTES**
> 1. A valid statement of reproducibility requires specification of the conditions changed.
> 2. The changed conditions may include: principle of measurement, method of measurement, observer, measuring instrument, reference standard, location, conditions of use and time.
> 3. Reproducibility may be expressed quantitatively in terms of the dispersion characteristics of the results.
> 4. Results here are usually understood to be corrected results.

Reproducibility includes both long and short-term random components and is therefore more representative of the variation actually experienced in day to day usage of the measuring instrument. In the experiments run at the Savannah River Site, only the time and associated environmental conditions varied. The same procedure, operator, test objects and techniques were used for all phases of the experiment. The uncertainty of most measuring and test equipment is affected by variations in the environment in which they are used and the degree to which the various parameters are controlled.

Figure 1 Mettler Toledo AX205 balance

It was decided a data collection experiment would be conducted to provide an empirical estimate of the uncertainty associated with measurements made with this balance at the Westinghouse Savannah River Technical Center.

Phase I - Experiment and Data Analysis

The first experiment involved using the Mettler

Toledo 5-place balance (Figure 1) purchased for pipette calibration work and research projects. Data needed to be collected for use in estimating uncertainty of the balance measurements. The balance variability and associated uncertainty directly impacts pipette calibration uncertainty budgets. The first experiment consisted of repeated measurements of a 100-gram standard, which was half of the capacity of the balance. A calibrated weight of 100.00151 +/- 0.00012 g @ k=2, was used. The measurements were taken once or more each working day as time allowed and spread over many days to provide good estimates of reproducibility. Replication "within" days as well as replication "between" days was desired to determine separate estimates of short-term (repeatability) and long-term random variation (reproducibility)[3]

It was decided that approximately 20 sets of 10 replicates would be performed to provide the data for statistical analysis. Each set consisted of 10 replicates that were made after an internal balance adjustment/calibration was performed. The experiment began in February 2001 in an office/lab environment with the usual office-type heating and cooling that had no humidity control. Because the operating environment was not tightly controlled, data on barometric pressure, temperature, and relative humidity were collected at time of weighing to determine their effect, if any, on the weighing variation.

The mean of the 10 measurements varied significantly from day to day. Also, the observed variation was significantly greater than the manufacturer's +/- 0.00003-gram tolerance for repeatability. The experiment continued until 22 sets of measurements were taken to provide adequate data for statistical evaluation. The measurement data and some summary statistics are shown in Table 1 below. As one can see, the variation in the daily average of 10 measurements varied as much as 0.34 milligrams, which is excessive for a 5-place AX205 balance. After contacting the distributor, the balance was returned to the manufacturer and replaced with the same type of balance. A thorough statistical analysis was completed on the measurement data to provide estimates of short and long-term random balance variation as originally planned. Analysis of the environmental data is presented below.

Date	1	2	3	4	5	6	7	8	9	10	Average	Std. Dev.
2/15/01	100.00145	100.00144	100.00148	100.00148	100.00150	100.00148	100.00148	100.00161	100.00177	100.00177	100.00155	0.00012669
2/15/01	100.00155	100.00157	100.00168	100.00152	100.00152	100.00154	100.00154	100.00156	100.00158	100.00156	100.00156	0.00004590
2/15/01	100.00159	100.00158	100.00158	100.00155	100.00156	100.00155	100.00156	100.00153	100.00156	100.00157	100.00156	0.00001757
2/22/01	100.00174	100.00169	100.00169	100.00148	100.00146	100.00144	100.00150	100.00150	100.00150	100.00154	100.00155	0.00010948
2/22/01	100.00151	100.00150	100.00148	100.00150	100.00145	100.00155	100.00147	100.00153	100.00146	100.00152	100.00150	0.00003199
2/26/01	100.00140	100.00139	100.00137	100.00136	100.00136	100.00137	100.00134	100.00134	100.00133	100.00133	100.00136	0.00002426
2/26/01	100.00161	100.00162	100.00162	100.00167	100.00162	100.00165	100.00162	100.00165	100.00163	100.00158	100.00163	0.00002495
2/27/01	100.00150	100.00149	100.00149	100.00147	100.00151	100.00144	100.00151	100.00148	100.00143	100.00145	100.00148	0.00002863
2/28/01	100.00164	100.00166	100.00164	100.00164	100.00160	100.00158	100.00160	100.00157	100.00159	100.00157	100.00161	0.00003316
3/01/01	100.00171	100.00169	100.00164	100.00171	100.00167	100.00164	100.00165	100.00156	100.00169	100.00164	100.00166	0.00004496
3/05/01	100.00157	100.00157	100.00154	100.00154	100.00144	100.00150	100.00140	100.00146	100.00143	100.00145	100.00149	0.00006202
3/06/01	100.00144	100.00140	100.00152	100.00145	100.00145	100.00147	100.00148	100.00143	100.00146	100.00143	100.00145	0.00003267
3/07/01	100.00152	100.00155	100.00155	100.00156	100.00160	100.00159	100.00157	100.00160	100.00160	100.00159	100.00157	0.00002751
3/07/01	100.00154	100.00155	100.00147	100.00149	100.00156	100.00148	100.00151	100.00149	100.00148	100.00152	100.00151	0.00003217
3/08/01	100.00156	100.00146	100.00148	100.00148	100.00143	100.00145	100.00146	100.00148	100.00145	100.00146	100.00147	0.00003510
3/12/01	100.00152	100.00159	100.00155	100.00153	100.00159	100.00155	100.00154	100.00156	100.00156	100.00157	100.00156	0.00002313
3/13/01	100.00143	100.00147	100.00149	100.00144	100.00143	100.00150	100.00139	100.00144	100.00148	100.00144	100.00145	0.00003352
3/13/01	100.00144	100.00139	100.00139	100.00144	100.00142	100.00149	100.00146	100.00147	100.00157	100.00152	100.00146	0.00005662
3/14/01	100.00135	100.00133	100.00135	100.00138	100.00132	100.00133	100.00130	100.00132	100.00128	100.00136	100.00133	0.00002938
3/18/01	100.00156	100.00156	100.00157	100.00161	100.00156	100.00159	100.00159	100.00162	100.00160	100.00164	100.00159	0.00002792
3/26/01	100.00150	100.00147	100.00146	100.00149	100.00147	100.00146	100.00145	100.00146	100.00144	100.00144	100.00146	0.00001962
3/27/01	100.00163	100.00166	100.00169	100.00170	100.00170	100.00166	100.00168	100.00168	100.00171	100.00166	100.00168	0.00002453
Grand Average = 100.001522 g												

Table 1 – Phase I Replicate Weight Measurements Data

Using analysis of variance (ANOVA) techniques available in Excel[R] spreadsheets[4], the data were separated into the two components, which are "within" days and "between" days. The "within" day component was calculated directly by the spreadsheet as a variance, which is shown in Figure 2 in bold type as 2.3636E-09 in the MS (mean square) column. The estimation of the "between" days component is slightly more difficult and must be calculated manually from the mean square value, MS (between), shown in the "between" days row. This value is known from statistical theory to

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Days	1.61305E-06	21	7.6811938E-08	32.5	8.33E-53	1.61
Within Days	4.6799E-07	198	2.3635838E-09			
Total	2.0810E-06	219				

MS(Between) = 7.68119E-08 = VarW + 10VarB

Therefore, VarB = { MS-VarW } / 10 = {7.681E-08 - 2.363E-09} / 10 = 7.444E-09

Total Var = VarW + VarB = 9.808E-09. Therefore **VarB is 75.9% of Total Var**.

Figure 2 – Phase 1 Between/Within Days Analysis

estimate the "within" days variance (Var W) plus *r* times the "between days" variance (Var B), where *r* is the number of "within" replicates. See Figure 2 for additional details in calculation of this estimate. Since *r* was 10, the "between" day variance estimate was calculated as 7.4448354E-09. Summing the "within" and "between" variances, the total random variance is 9.8084192E-09. This is equivalent to a standard deviation of 0.00009904 gram or approximately 0.1 milligram. This estimate is considerably larger than the manufacturer's repeatability estimate, which is 3 scale divisions or 0.03 mg.

From the total random variance and the "between" day component, the "between" day variance was calculated as 75.9% of the total. It should also be noted that an estimate of total variation would be a serious underestimate if one only considered "within" day variation. The "between" day component is 3 times as large and is primarily responsible for the large variation experienced with the balance. As a result, future experimental effort with the new balance was directed towards minimizing the "between" day variance.

Phase II - Experiment and Data Analysis

After receiving the new AX205 balance in early April 2001, it was decided to perform the experiment on the new balance exactly as had been done previously so that proper statistical estimates could be made. Data were collected from April 9 through mid-May and are shown in Table 2 below.

From the beginning, the new balance appeared to be much more stable and have less inherent variability. Data collection continued until May 16 when 22 sets of data had been generated. The smaller magnitude of variability was confirmed after calculating the "within" day variations that are shown in Table 2. As was done previously, a complete statistical treatment was performed on the data, and the analysis is shown in Figure 3.

Date	1	2	3	4	5	6	7	8	9	10	Average	Std. Dev.
4/9/01	100.00141	100.00141	100.00139	100.00140	100.00141	100.00139	100.00143	100.00142	100.00138	100.00140	100.00140	0.00001505
4/10/01	100.00136	100.00135	100.00134	100.00137	100.00138	100.00139	100.00135	100.00138	100.00142	100.00141	100.00138	0.00002634
4/10/01	100.00147	100.00147	100.00147	100.00148	100.00146	100.00145	100.00146	100.00145	100.00145	100.00147	100.00146	0.00001049
4/10/01	100.00144	100.00142	100.00143	100.00143	100.00142	100.00142	100.00141	100.00138	100.00139	100.00136	100.00141	0.00002540
4/11/01	100.00153	100.00152	100.00153	100.00152	100.00152	100.00152	100.00153	100.00155	100.00152	100.00154	100.00153	0.00001025
4/11/01	100.00151	100.00149	100.00149	100.00149	100.00147	100.00148	100.00150	100.00147	100.00148	100.00147	100.00149	0.00001340
4/12/01	100.00153	100.00151	100.00149	100.00147	100.00148	100.00147	100.00147	100.00147	100.00147	100.00148	100.00148	0.00002050
4/12/01	100.00145	100.00142	100.00145	100.00144	100.00148	100.00141	100.0014	100.00138	100.00146	100.00145	100.00143	0.00003062
4/16/01	100.00144	100.00143	100.00143	100.00142	100.00140	100.00138	100.00141	100.00137	100.00136	100.00137	100.00140	0.00002927
4/17/01	100.00140	100.00139	100.00138	100.00139	100.00138	100.00140	100.00140	100.00142	100.00139	100.00140	100.00140	0.00001179
4/23/21	100.00148	100.00148	100.00147	100.00149	100.00148	100.00151	100.00149	100.00149	100.00147	100.00149	100.00149	0.00001179
4/25/01	100.00142	100.00143	100.00145	100.00146	100.00145	100.00146	100.00146	100.00147	100.00146	100.00147	100.00145	0.00001633
4/26/01	100.00152	100.00152	100.00149	100.00144	100.00145	100.00142	100.00145	100.00143	100.00141	100.00145	100.00146	0.00003913
4/30/01	100.00144	100.00145	100.00144	100.00148	100.00150	100.00141	100.00145	100.00144	100.00143	100.00146	100.00145	0.00002546
5/01/01	100.00145	100.00132	100.00145	100.00142	100.00140	100.00137	100.00140	100.00142	100.00142	100.00136	100.00140	0.00004093
5/02/01	100.00128	100.00132	100.00129	100.00129	100.00132	100.00128	100.00131	100.00131	100.00132	100.00128	100.00130	0.00001767
5/03/01	100.00147	100.00152	100.00150	100.00146	100.00143	100.00146	100.00147	100.00147	100.00145	100.00149	100.00147	0.00002568
5/07/01	100.00140	100.00141	100.00138	100.00140	100.00140	100.00139	100.00140	100.00144	100.00138	100.00138	100.00140	0.00001812
5/08/01	100.00140	100.00142	100.00138	100.00139	100.00138	100.00139	100.00137	100.00138	100.00137	100.00137	100.00139	0.00001588
5/14/01	100.00150	100.00151	100.00144	100.00143	100.00147	100.00144	100.00137	100.00143	100.00145	100.00142	100.00145	0.00004029
5/15/01	100.00144	100.00145	100.00145	100.00144	100.00145	100.00144	100.00145	100.00143	100.00144	100.00145	100.00144	0.00000696
5/16/01	100.00154	100.00148	100.00150	100.00154	100.00153	100.00155	100.00154	100.00155	100.00153	100.00153	100.00153	0.00002246
Grand Average = 100.001436 g												

Table 2 – Phase II Replicate Weight Measurements Data

The analysis indicated the new balance was superior to the old in both the "within" and "between" day variation. The "within" day variation improved from a variance of 2.3636E-09 to 5.6209E-10 while the "between" day variation improved from a variance 7.4448354E-09 to 2.78211E-09. Summing the components as before, gave a total random variance of 3.34419E-09. This is equivalent to a standard deviation of 0.000058 gram or approximately 0.06 mg, which compares to 0.1 mg with the first balance.

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Days	5.96046E-07	21	2.83832E-08	50.5	8.03E-68	1.61
Within Days	1.11293E-07	198	5.62086E-10			
Total	7.07339E-06	219				

MS(Between Days) = 2.83832E-08 = VarW + 10VarB

Therefore, VarB = { MS-VarW } /10 = { 2.83832E-08 - 5.62086E-10 } / 10 = 2.78211E-09

Total Var = VarW + VarB = 3.344E-09. **Therefore VarB is 83.2% of Total Var**.

Figure 3 – Phase II Between/Within Days Analysis

Phase III - Experiment and Data Analysis

Even though the total variation was improved over the first balance, a calculation of "between" day contribution to the total variation indicated it was still the largest contributor at 83%. As a result, additional effort was undertaken to further improve the day to day balance variation. The balance was placed on a marble slab, a new power supply was obtained from the manufacturer and a field service technician came to the laboratory to perform his own diagnostics. While in the laboratory, the technician re-calibrated the internal weights with OIML Class E2 mass standards. After these modifications, additional measurements were taken for 22 days, beginning in early August and ending in mid-September. These measurements are shown in Table 3.

The same statistical analysis was performed on the data as was done previously. The analysis is shown in Figure 4. The “between”, “within” and total estimates of random variance were slightly larger than previous, but the differences were determined not to be statistically significant. From

Date	1	2	3	4	5	6	7	8	9	10	Average	Std. Dev.
8/07/01	100.00151	100.00149	100.00152	100.00154	100.00153	100.00151	100.00149	100.00151	100.00148	100.00149	100.00151	0.0000195
8/07/01	100.00150	100.00148	100.00148	100.00150	100.00150	100.00148	100.00149	100.00148	100.00147	100.00143	100.00148	0.0000208
8/08/01	100.00150	100.00145	100.00145	100.00142	100.00141	100.00142	100.00144	100.00144	100.00144	100.00145	100.00144	0.0000249
8/08/01	100.00144	100.00143	100.00143	100.00140	100.00141	100.00142	100.00141	100.00141	100.00142	100.00144	100.00142	0.0000138
8/09/01	100.00136	100.00138	100.00138	100.00138	100.00138	100.00138	100.00138	100.00138	100.00138	100.00138	100.00138	0.0000058
8/09/01	100.00142	100.00140	100.00139	100.00141	100.00140	100.00139	100.00140	100.00141	100.00140	100.00140	100.00140	0.0000093
8/13/01	100.00148	100.00147	100.00148	100.00149	100.00149	100.00146	100.00146	100.00145	100.00141	100.00139	100.00146	0.0000336
8/14/01	100.00150	100.00148	100.00146	100.00147	100.00147	100.00146	100.00145	100.00143	100.00144	100.00141	100.00146	0.0000259
8/15/01	100.00141	100.00142	100.00142	100.00141	100.00143	100.00138	100.00135	100.00133	100.00134	100.00134	100.00138	0.0000394
8/20/01	100.00152	100.00150	100.00152	100.00152	100.00152	100.00148	100.00146	100.00144	100.00140	100.00141	100.00148	0.0000471
8/27/01	100.00159	100.00157	100.00157	100.00154	100.00154	100.00155	100.00154	100.00154	100.00155	100.00153	100.00155	0.0000189
8/28/01	100.00148	100.00148	100.00148	100.00149	100.00146	100.00140	100.00148	100.00146	100.00144	100.00143	100.00146	0.0000287
8/29/01	100.00164	100.00164	100.00162	100.00162	100.00163	100.00162	100.00161	100.00161	100.00160	100.00160	100.00162	0.0000144
8/30/01	100.00157	100.00156	100.00157	100.00159	100.00157	100.00160	100.00158	100.00160	100.00157	100.00156	100.00158	0.0000149
9/04/01	100.00159	100.00161	100.00153	100.00152	100.00152	100.00151	100.00151	100.00149	100.00150	100.00152	100.00153	0.0000389
9/05/01	100.00149	100.00151	100.00152	100.00149	100.00150	100.00153	100.00149	100.00149	100.00151	100.00150	100.00150	0.0000144
9/06/01	100.00147	100.00147	100.00146	100.00147	100.00147	100.00148	100.00147	100.00147	100.00145	100.00146	100.00147	0.0000082
9/12/01	100.00154	100.00150	100.00155	100.00154	100.00151	100.00155	100.00153	100.00152	100.00152	100.00151	100.00153	0.0000176
9/13/01	100.00159	100.00156	100.00153	100.00155	100.00154	100.00157	100.00157	100.00156	100.00152	100.00155	100.00155	0.0000206
9/17/01	100.00150	100.00149	100.00143	100.00146	100.00147	100.00142	100.00142	100.00142	100.00142	100.00142	100.00145	0.0000320
9/18/01	100.00150	100.00150	100.00147	100.00148	100.00149	100.00150	100.00147	100.00146	100.00146	100.00146	100.00148	0.0000173
9/19/01	100.00155	100.00152	100.00159	100.00155	100.00155	100.00152	100.00146	100.00147	100.00152	100.00148	100.00152	0.0000412
Grand Average = 100.001484 g												

Table 3 – Phase III Replicate Weight Measurements Data

the total calculated variance of 4.4178355E-.09, the total random standard deviation was calculated as 0.000066 gram or approximately 0.07 milligram. Apparently, the modifications made to the balance and environment had little effect on the random variation of the balance. Having made no apparent improvement in the random variation, it was concluded a standard deviation of 0.06 to 0.07 milligram is the best that can accomplish with the balance considering the uncontrolled environmental conditions. This uncertainty is approximately twice the “repeatability” value stated by the manufacturer. Additional effects of the environment are discussed below.

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Days	8.0373E-07	21	3.83E-08	58.3	5.13E-73	1.61
Within Days	1.2992E-07	198	6.56159E-10			
Total	9.3365E-07	219				

MS(Between Days) = 3.827E-08 = VarW + 10VarB
Therefore, VarB = { MS-Var(within) } /10 = { 3.82729E-08 - 6.56159ZE-10 } / 10 = 3.76168E-09
Total Var = VarW + VarB = 4.41783E-09. **Therefore VarB is 85.1% of Total Var**.

Figure 4 – Phase III Between/Within Days Analysis

Environmental Effects on Variation

Day to day variation was experienced in all phases of the experiment. A good summary of the weight variation over time is a plot of daily averages as shown in Figure 5 below. The effect of environmental changes is most likely the cause of the day to day weight variation seen in the data. Table 4 is a summary of the environmental parameters that were measured during Phase

I, II and III. They were temperature in centigrade (Temp C.), barometric pressure in hectopascals (BP) and percent relative humidity (%Humidity). The average gram quantity and standard deviation are shown.

As one can see, there was some variation in all the environmental factors measured. The AX205 model balance has a temperature compensation feature that should have corrected for any temperature variation of 1-degree C. The barometric pressure differences will cause air buoyancy effects, but their significance should be at a lesser order of magnitude. However, humidity could possibly affect the weights by moisture gain or loss on the surface of the internal calibration weights or the surface of the nominal 100-gram test weight. Figure 6 shows that the highest humidity standard variations were during Phase I, least in Phase II and slightly more in Phase III. Using an F-test on the corresponding variances, the humidity standard deviations were found to be statistically significant from one another with 95% confidence.

Environmental Factors				
		Phase I	Phase II	Phase III
Weight (grams)	Avg. Std. Dev.	100.00152 0.000099	100.00144 0.000058	100.00148 0.000066
Temp C	Avg. Std. Dev.	23.09 0.95	22.67 0.89	22.40 0.75
BP(mm)	Avg. Std. Dev.	1005.1 4.25	1007.25 4.65	1004.72 2.51
% Humidity	Avg. Std. Dev.	36.7 8.0	36.5 3.2	45.8 5.4

Table 4 – Environmental Data Summary

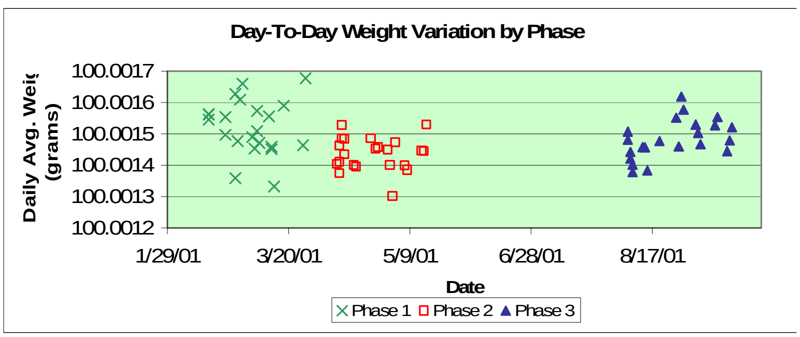


Figure 5 - Day-to-Day Weight Variation for All Phases

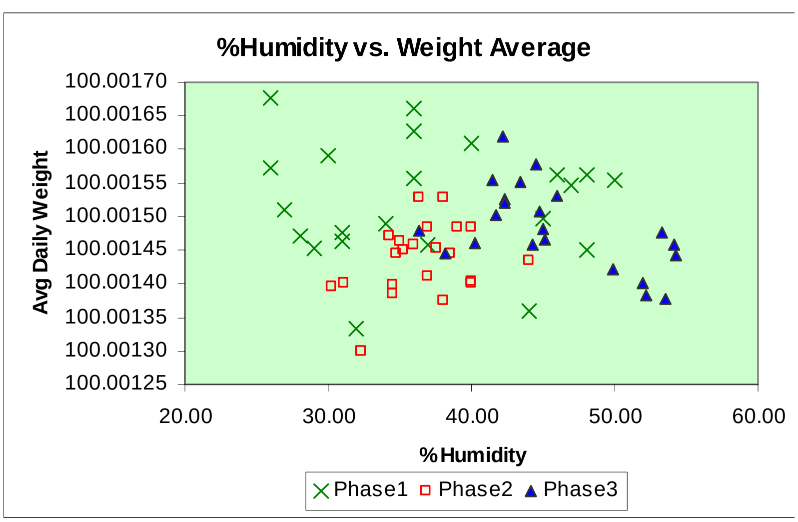


Figure 6 - Day-to-Day Weight vrs. Humidity Variation

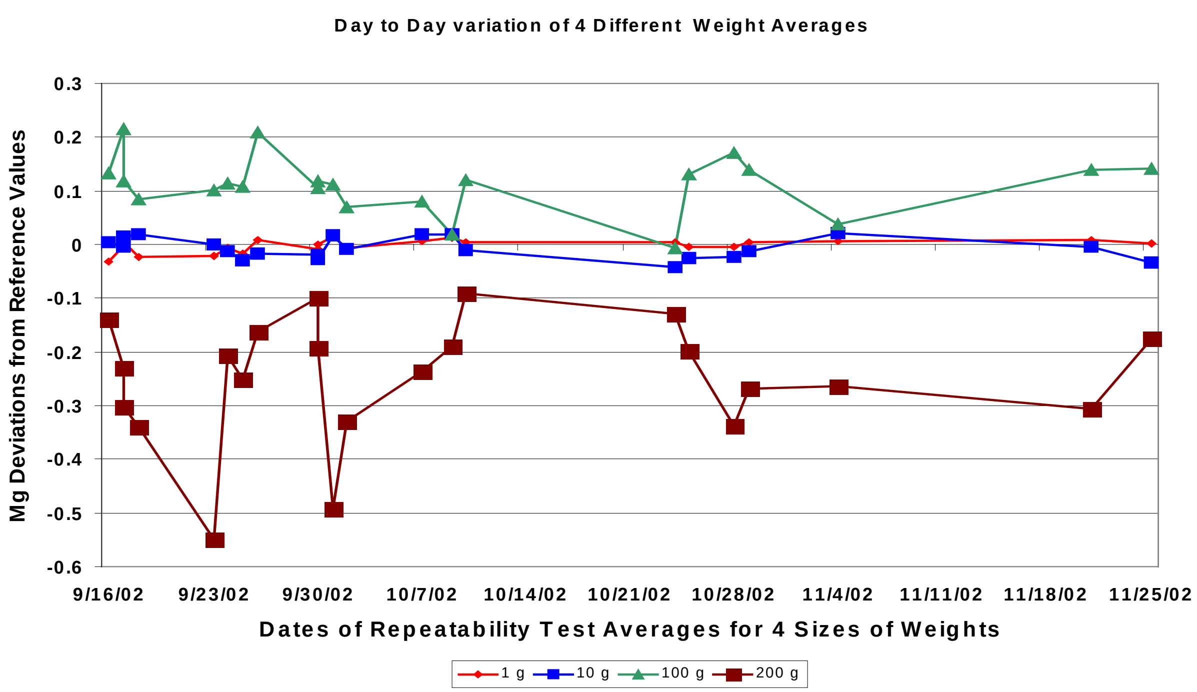
The day-to-day weight variations follow the same pattern in magnitude (See Figure 5 or Table 4), and suggest that the humidity variation was a large contributor to the day-to-day variability.

Bias Considerations

So far this discussion has centered only on the random uncertainty of the balance. Now the systematic error or bias will be considered. Of most interest is a comparison of the second balance before (Phase II) and after calibration (Phase III). The grand averages for these phases from Tables 2 and 3 are 100.001436 and 100.001484 respectively, from which we calculate a difference of 0.0000473 gram or 0.05 milligram. Using a t-test, this difference is significant which indicates a shift in the average weight measurements, which is possibly due to the recalibration by the Mettler technician. Also, the Phase III average is nearer to the certified weight value of 100.00151 grams than before the calibration.

Phase IV - Repeatability tests over the full range of the balance

Since this balance was purchased to calibrate pipettes, it was decided to do the repeatability test again with standards above and below the 100 g check standard used in the first phases. This experiment used the same "repeatability" test of 10 measurements with 1, 10, 100 and 200 g test weights. The same procedure operator, procedure and balance were used to make an additional 22 sets of measurements. The environmental conditions were recorded after the balance was adjusted / calibrated. First the 200 g weight was measured 10 consecutive times and the zero checked. The balance was then zeroed and the test repeated for the next largest weight down to the 1 g weight. After the 10 g weight was tested, it was left on the pan and the balance tared to zero. The 1 g weight was placed on top of the 10 g weight and weighed. It was removed, replaced and weighed again. The process was repeated until 10 measurements had been collected. The averages and standard deviations were determined for each set of data. Averages of all weights varied from run to run. Figure 7 is a plot of each repeatability test average value minus its respective reference value. The plots of each weight graphically illustrate the difference in magnitude of the variations for the different sizes of weights. The systematic error is reflected in the negative direction for the 200 g weight, which was an OIML class E2 weight purchased from and calibrated by a major balance manufacturer for use as a check standard. The 100 g weight was the same weight used in the previous tests and shows an average positive bias. The magnitude of its variations is half that of the 200 g weight and much larger than the magnitude of the 1 and 10 g weights.



.Figure 7 - Plot of the Repeatability Test Average Weights of Phase IV Experiment.

Table 5 is a summary of the ANVO tests for the Phase IV experiment. It is interesting to note that the "between" days variation accounts for over 80% variation for the larger sizes of weights. This is probably due to the shifts in the slope of the line between the span and zero. The variation is proportional to the magnitude of the weights. Note the total standard deviation (reproducibility)

Weight(gms)	Ave Bias	Betw. Days SD.	Within Days SD	Total Std. Dev.	Betwn. %
1	-0.213	1.1661	1.070	1.582	54
10	-0.714	1.8782	1.484	2.394	62
100	11.348	5.1959	1.884	5.527	88
200	-24.732	11.7543	2.572	12.032	95

Table 5 - Summary of Analyses of Variance for 4 Weights in the Phase IV Experiment

10	0.07	-0.07	-0.24	0.20
100	-0.35	-0.24	0.07	-0.01
200	-0.19	-0.07	0.20	0.39

Table 6 - Correlations to Environmental Conditions

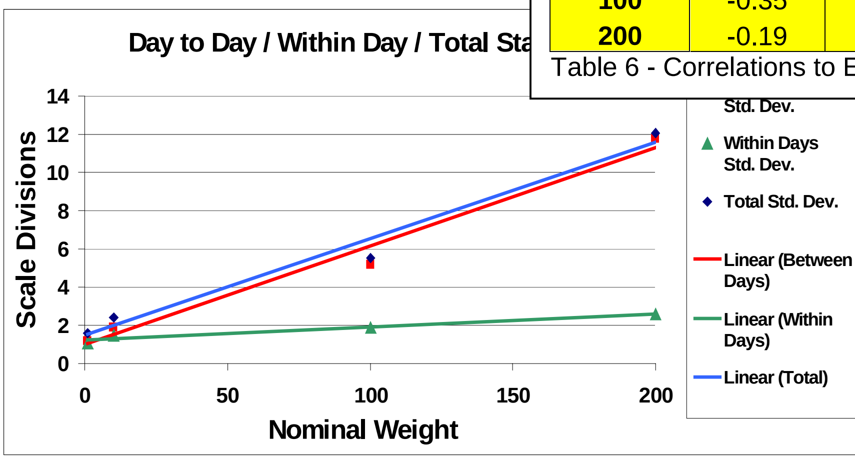


Figure 8 - Relationship of Between & Within Variation of Weights

estimates for the 1 and 10 g weights are within manufacturer's specifications.

Figure 8 illustrates the increasing magnitude of the "between" days variation as the weights increase in size. It clearly illustrates that the majority of the error comes from the reproducibility variation due to the changing environmental conditions. This experiment did not test the 25% range of the weighing scale, so a Phase V experiment was conducted to collect data to adequately quantify the reproducibility of the 5 place analytical balance in the SRS office environment.

The correlation coefficients of the major environmental factors are given in Table 6 for each weight tested in Phase IV. Only the room temperature appears to have a large correlation in the repeatability tests for the 1 g weight. This is shown in Figure 9. The chart shows the range of temperatures in the office varied by as much as 5º C. The correlations for the other parameters and weights were not as significant.

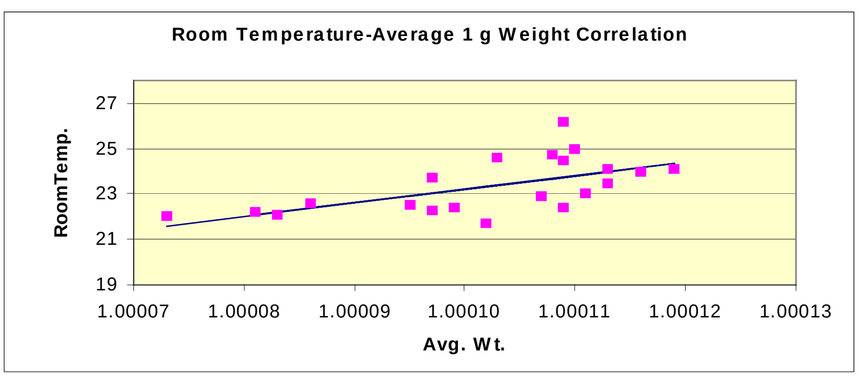


Figure 9 -Repeatability Averages vrs. Temperature Correlation

Phase V - Second repeatability test over the full range of the balance

The full range repeatability experiment was conducted with a new weight set of Troemner Ultra Class™ 1, 10, 50,100 and 200 g mass standards. This set of weights was packaged especially for quality checks on 200 g balances. This experiment provided repeatability information at 25% of the operating range of the balance. Figure 10 summarizes the normalized average values of the 17 runs that were made for this test. Each repeatability test was normalized by subtracting the conventional mass reference value from the average. The 50 g weight total standard deviation was closer to that of the 100 g weight than the 10 g weight. The "between" days variation was again the dominate source of variation in the reproducibility of these measurements. Summaries of the ANOVA for each of the weights tested are shown in Table 7. Again the "reproducibility" standard deviation for all of the measurements made during the experiment for the 1 and 10 gram weights were within the manufacturers +/-0.03 mg specification for repeatability. The total standard deviations of the 170 measurements for 50, 100 and 200 g weights exceeded it.

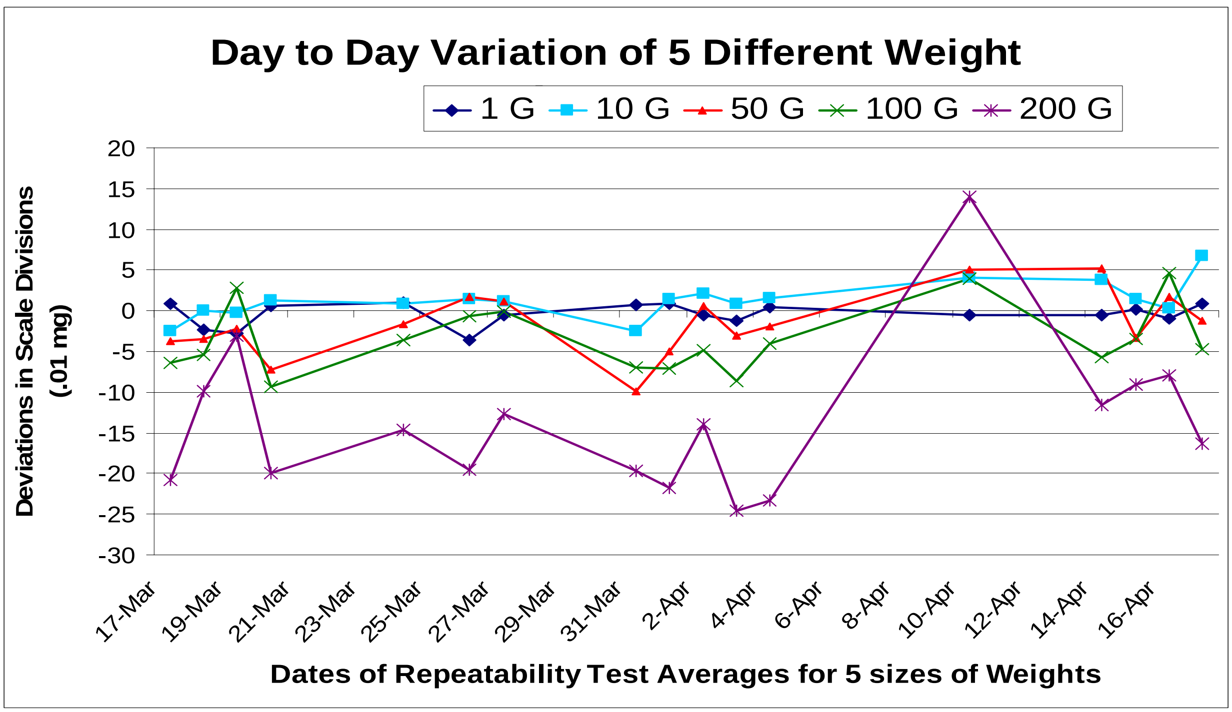


Figure 10 - Plot of the Normalized Repeatability Test Average Weights of Phase V Experiment

Weight(gms)	Ave Bias	Betw. Days SD. St	Within Days SD	Total Std. Dev.	Betwn. %
1	-0.471	1.3380	1.066	1.711	61
10	1.276	2.2029	1.103	2.464	80
50	-1.603	3.9490	0.918	4.054	95
100	-3.519	4.1906	1.616	4.491	87
200	-13.800	9.3199	2.170	9.569	95

Table 7 - Summary of Analyses of Variance for 5 Weights in the Phase V Experiment

The total standard deviations observed during this experiment, based on only 17 runs, were smaller for the larger weights, than those observed in the other experiments that had 22 runs and were conducted over a longer period of time.

Figure 11 is much like Figure 8 in graphically illustrating the significant difference between the "within" (repeatability) variation and the "between" (reproducibility) days variation. The conclusion that can be drawn for these experiments is the weighing system, including the environmental conditions has significant "between" times variation that adds to the uncertainty of

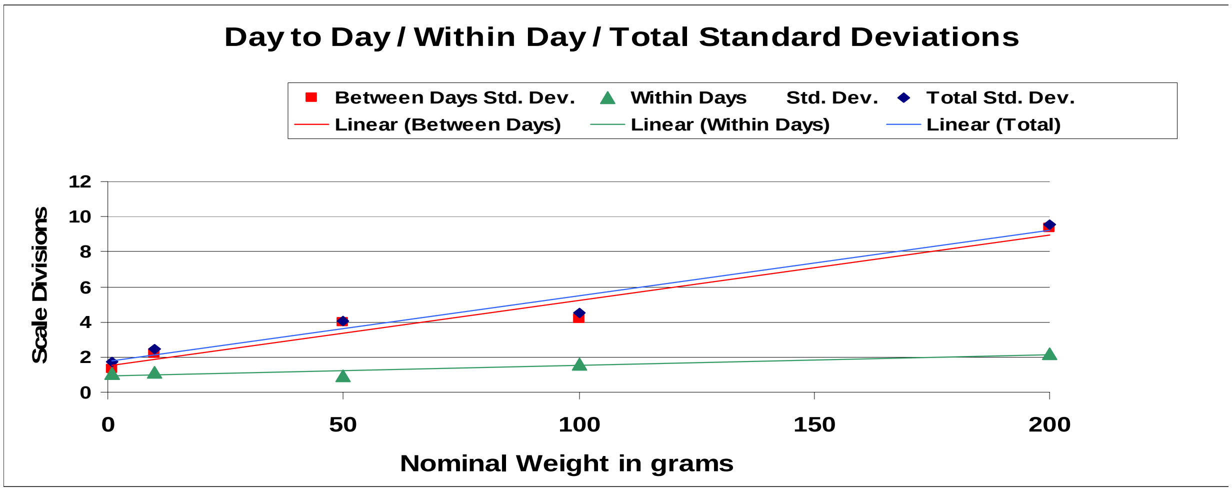


Figure 11- Phase V experiment relationships of the between and within days variations

knowing the true value of measurements made in direct reading of the balance. For balance users to adequately determine the uncertainty of their measurements, they should routinely measure an artifact that mimics the mass of the material they will routinely weigh and use the standard deviation of all the measurements collected over time to estimate the measurement process uncertainty. The total standard deviation should be used in the propagation of the uncertainty estimate.

Linearity Tests

Repeatability is one of the 3 tests that are conducted for confirming the performance of the balance meets manufacturer's specifications. Linearity is also tested. Most of the balance calibration industry tests linearity by weighing a test objects that is ~ 25% of the capacity of the balance. For the 200 g balance, the 50 g weight used in the phase V "repeatability" test was used in the linearity test because its reference value was well known. The reference value of the test object used by the industry does not need to be known. The basic procedure they use involves weighing the test object at the 25%, 50%, 75% and 100% scale loadings positions. This is done by taring the balance, first with an empty pan, next with tare weights equal to 25%, 50% and 75% of the balance capacity and placing the test item on top of the weigh that has been tared to give a zero display. The average of the 4 readings is subsequently subtracted from each of the measurements and the differences compared to the tolerance. For the AX205 balance linearity tolerance is +/- 0.1 mg, or 10 scale divisions.

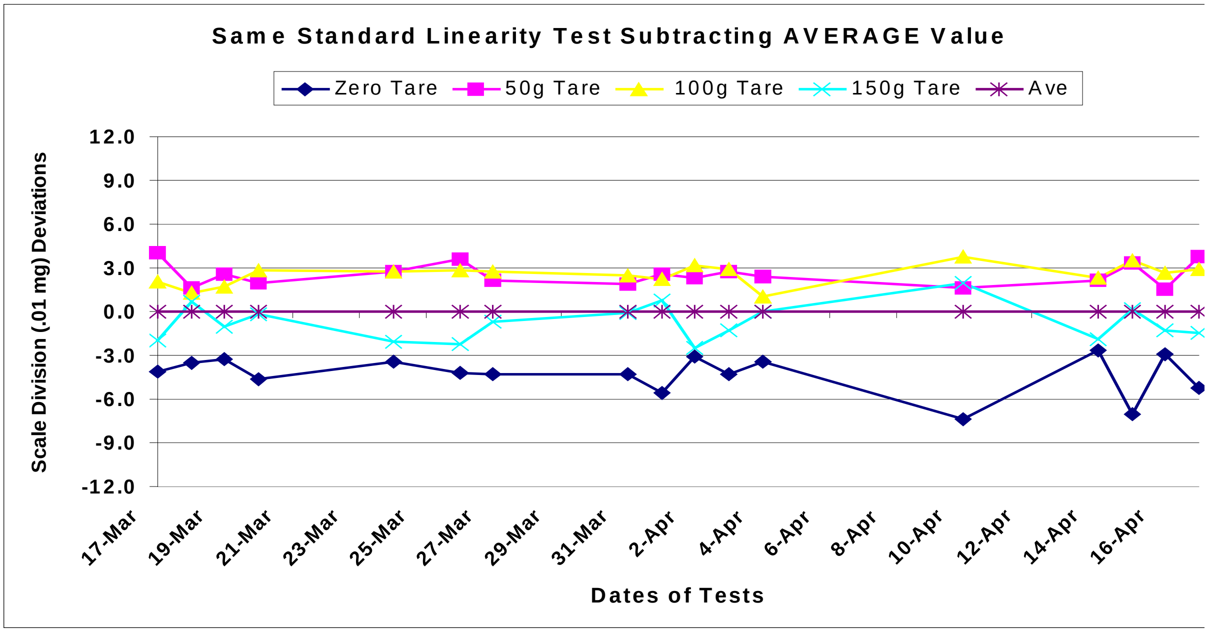


Figure 12 –Summary of Industry Linearity Test Tests Run 17 Times Using a 50 g Weight

As part of the Phase V experiment, the 50 g weight was used to test the balance at the 25%, 50%, 75% and 100% scale loadings positions, with appropriate tares. Instead of one measurement at each location, the experiment involved 5 sets of readings. The 5 sets of measurements were conducted 17 times. The average results for each part of the weighing range are plotted in Figure 12 after subtracting average value. The day to day variation is apparent in the graph. The non-linearity is fairly consistent, regardless of the shifts in the averages for the day of test. The average maximum non-linearity is 6.8 scale divisions as compared to the manufacturer's 10 scale division tolerance.

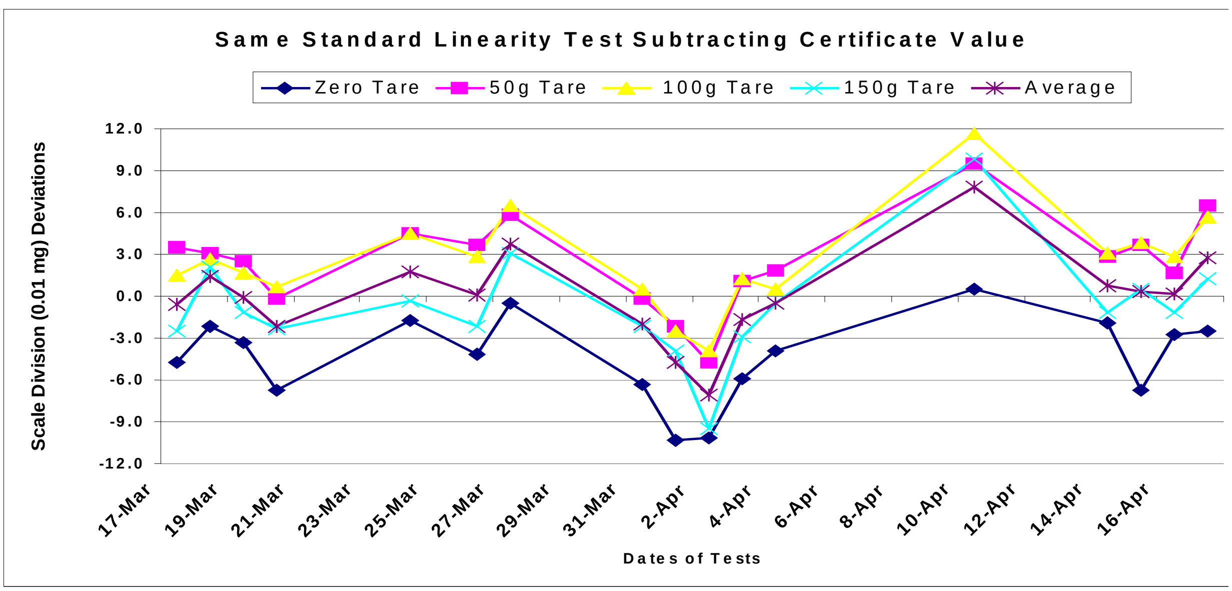


Figure 13 – Summary of same Linearity Test Run 17 Times after Subtracting Reference Value

Table 7 summarizes the average measurements of the 50 gram weight at the 0, 50,100 and 150 g locations. It shows the bias two ways, first by subtracting the reference value and second by subtracting the average value. The net non-linearity estimates are the same. It is largest between the

Tare =	0 g	50 g	100 g	150 g
Average =	49.999990	50.000059	50.000059	50.00002(
Bias (scale divisions)	-3.8	3.0	3.0	-0.3
Stdev = (same)	4.3	3.7	3.8	4.3
Industry Bias = (same)	-4.3	2.5	2.5	-0.8

Table 7 Summary statistics of 85 measurements with each tare

0 and 50 g tare. Both methods give the same 6.8 scale divisions. The industry method of subtracting the average is shown in Figure 12. It does not show the day to day variability that is removed by subtracting the daily average. All of the readings fall well within the 10 scale division tolerance.

Figure 13 shows the same repeatability test averages of 5 individual linearity tests after the reference value is subtracted. This chart shows the effects of random variation due to changes in room temperature, balance temperature, humidity and barometric pressure. The experiments reveal short term systematic shifts in the measurements made using stable artifacts. The overall variations of the readings at the 4 positions on the range of the balance vary from -10 to +12 scale divisions. The net effect of the short term systematic errors is an expanded estimate of the random variation of the measurement system. When calculating the uncertainty estimate for measurements made with a balance, it is wise to use the standard deviation of the measurements of a check stand in a control program, as is often done in regulated laboratories.

Example of Measurement Uncertainty Estimates Using QA Program Data

One laboratory uses several balances for measurements of small quantities of material. They routinely verify the balances are performing within specific control limits. They have two weight sets that they switch out periodically. Uncertainty estimates were calculated by using the

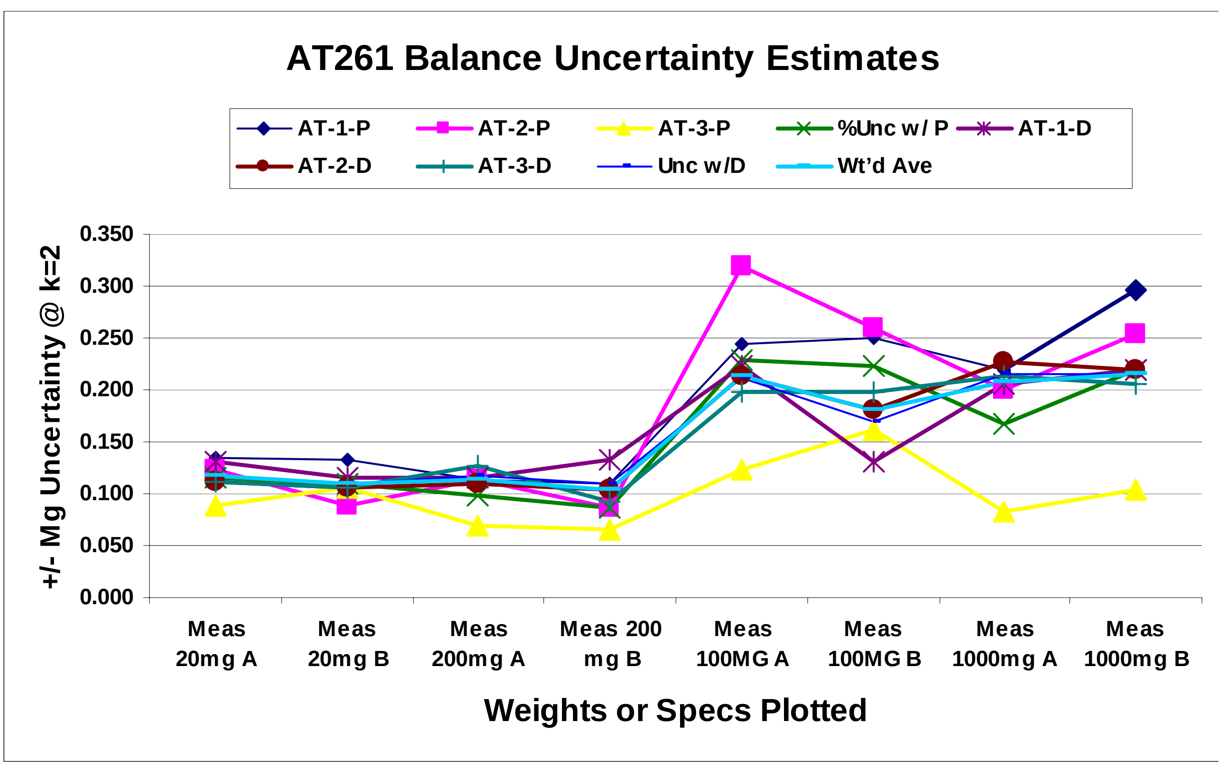


Figure 14 - Uncertainty Estimates for 3 UMT Balances Using 2 Different Weights

"reproducibility" standard deviation for all of the measurements made with each check standard. The uncertainties are plotted in Figure 14 for Mettler Toledo AT261 balances. The uncertainties are different for each weight that is routinely measured twice in testing the daily QA test. This chart illustrates the uncertainty in estimating uncertainty of measurements. The first four points on the chart are based on making measurements to the nearest 0.01 mg using the 5 decimal place display. The next 4 points are based on using the 4 decimal place display, or reading to 0.1 mg. The balance measurement uncertainty estimates ranged from 6 scale divisions to 13 for the 5 place range. The four place range uncertainty estimates varied from 0.8 to 3.2 scale divisions for the two different sizes of check weights. The uncertainty calculations reveal more variation in using the 5 place range of the balance in the laboratory environment. There were significant differences in the variation of the different balances also. This laboratory controls the temperature at 21.5 +/- 1° C and humidity at 55 +/- 5%. This chart illustrates the uncertainty in estimating uncertainty of measurements. Change a variable and the uncertainty estimate changes also. It is up to the user to study the measurement system and to determine which parameters have the most variation in them. Then decide if an average uncertainty should be used that truly captures the "reproducibility" of the various parameters of the measurement system.

Conclusions

These experiments demonstrate the need for determining the variation of the user's balance in their environment to determine a Type A estimate of their measurement process standard deviation. This can be done by designing an experiment that will capture the major variables in the reproducibility of measurements. The described experiments helped identify a balance that had a large reproducibility standard deviation that exceeded the manufacturer's repeatability tolerance by ~ 3 times. Subsequent experiments indicated the reproducibility at the author's facility provided standard deviations that were twice the tolerance, using a test weight equal to half the capacity of the balance.

Testing the balance over the full operating range provided evidence that the magnitude of the reproducibility was proportional to the mass of the test weight. The larger test objects have the greater "between" days variation. Therefore, uncertainty estimates should be developed for range of the balance that will be used for measurements.

Finally, the reproducibility in the bottom 5% of the balance range was within manufacturer's tolerance for repeatability, even through over half of the variation was due to the "between" days variation. These experiments illustrated the difference between "repeatability" and "reproducibility" variation in balance measurements. One should not assume there is one balance standard deviation that can be used in estimating the uncertainty of weight measurements.

References

[1] The information contained in this article was developed during work under Contract No. DE-AC09-96SR18500 with the U. S. Department of Energy. To be published as a WSRC-MS-2003-00322 document.
[2] International Vocabulary of Basic and General Terms in Metrology, Second Edition, 1993, ISBN 92-67-01075-1, International Organization for Standardization, printed in Switzerland.p.36.
[3] Box, G. E. P., Hunter, W. G., and Hunter, J. S., *Statistics for Experimenters*, John Wiley & Sons, Inc., 1978
[4] Walpole, R. E., and Myers, Raymond H., *Probability and Statistics for Engineers and Scientists*, Macmillan Publishing Co., Inc., 1972.