

What TUR Do You Really Need? Putting Statistical Theory into Practice

Warren Wong
Fluke Corporation

The ageless question in metrology is, “What Test Uncertainty Ratio (TUR) is required to properly calibrate an instrument?” Numerous papers on guardbanding and statistical techniques have been published to address this question. Most recently, these discussions have focused on the very conservative “in-tolerance” conditions proposed by ISO Guide 25, draft 5. Although ISO Guide 25 is being replaced by ISO 17025, and the present draft (DIS 17025) does not have the strict guardbanding requirements, the draft 5 proposal presented a perspective that has merit.^{[1][2]} The philosophies of other guardbanding methods have also been discussed in detail, especially by Dave Deaver of the Fluke Corporation.^{[3][4][5]}

This paper applies two different guardbanding methods to a real world example. The verification procedure for a Fluke 8842A system DMM using a Fluke 5520A Multi-Product Calibrator is performed numerous times. Two different guardbands are applied to those measurements with less than a 4:1 TUR and the yields are analyzed. Although this paper does not emphatically answer the question about what TUR is needed, or which guardband method to use, it should give insight into how you can apply these techniques to your own quality system.

The Experiment

Six instruments, using three different calibrators, were verified a total of 30 times. Calibrations were performed under various environmental conditions; with the intention of putting bias into the system, as well as changing the distribution of measurements. From earlier literature, guardbanding gives the best yields if the Unit-Under-Test (UUT) and calibrator have no biases, are normally distributed (i.e. Gaussian), and are “typically” better than their specifications.^[3] At Fluke, years of manufacturing experience with this UUT indicated that its normal distribution and lack of bias would have made the results of this experiment less interesting. By subjecting the UUT and calibrator to different temperatures, from 0°C to 50°C, the yields became distorted. Furthermore, this simulates results with a less than ideal UUT.

For this study, to maintain consistency with previous guardbanding papers, and to be in line with the accepted practice of accredited laboratories, the uncertainty specifications of the calibrator, normally stated by Fluke with a 99% confidence level (i.e. 2.58σ), was re-calculated to 2σ or 95% confidence. No assumption was made concerning the confidence level of the UUT. Hence, its specifications were accepted as is.

There are 130 verification points required for this UUT, with 10 points that have TURs less than 4:1 when compared against its 90 day specifications. These points are shown in Table 1.

Measurement Point	TUR
0 mV, 20 mV DC range	3.87
0 mV, 200 mV DC range	3.87
100 mV, 200 mV DC range	3.61
100 V, 200 V DC range	3.91
1000 V, 1000 V DC range	3.91
1 A, 2 A DC slow	2.84
1 A, 2 A DC medium	2.97
1 A, 2 A DC fast	3.87
100 Ω, 200 Ω range, 4-Wire	3.94
100 Ω, 200 Ω range, 2-Wire	3.94

Table 1: There are 10 points with < 4:1 TUR.

Guardband Philosophies

Calibration is about risk management. If measurement uncertainties are too high (or TURs are too low), what is the risk of performing a bad calibration?

Agy/Wong gave a guardbanding proposal based on ISO Guide 25 to ensure that a *given* measurement has a 95% probability of actually being within specification. Deaver’s earlier papers talk about consumer and producer risks when calibrating a population of instruments. He discussed guardbanding methods as a way to manage this risk.^{[3][4]} His paper at the 1998 NCSL considered more of the ISO Guide 25 perspective, calculating the probability the true value of a *given* measurement is within the specification limits.

Setting a guardband is nothing more than establishing a new test limit that determines when a measurement is in-tolerance. The Agy/Wong approach (hereafter called the ISO method) re-computes a test limit such that for any one given measurement, there is a 95% probability that a measurement is in-tolerance. This approach assumes no priori knowledge of the UUT. The 1998 Deaver proposal says that for normally distributed UUT and TURs between 1.5:1 and 4:1, a simple 80% guardband would give equal to or better than a 95% probability a measurement is in-tolerance. His earlier papers calculated test limits such that, in a population of instruments, the consumer risk (i.e. the risk of declaring a bad measurement good) was the same as a 4:1 TUR.

To further explain, in this study the 1A measurement, “slow” reading rate, has a TUR of 2.84:1. Deaver’s 1998 proposed guardband of 80% gives a better than 95% probability the UUT is in-tolerance if the distribution is normal.^[5] The ISO method uses a guardband that is 71% of the UUT specification, but assumes no knowledge of the UUT. If a reading falls inside that 71% window, there is at least a 95% probability that measurement is good. Deaver’s earlier papers took the

perspective of the manufacturer of the UUT. If an 80% guardband is taken, there is a consumer risk of 0.2% and a producer risk of 9%. In other words, for a population of instruments, there is 0.2% chance of declaring the 1 A measurement good when it really is bad; and a 9% chance of falsely declaring a good 1A reading.^[3] In a practical sense, what does all of this mean?

Applying the Guardbands

Thirty calibrations were performed on the UUTs. Guardbands, or new test limits, were applied for all TURs < 4:1. In the 1998 Deaver approach, this is accomplished by taking 80% of the UUT specifications (i.e. If the measured value is within 80% of the UUT specification, then it is declared in-tolerance). The ISO method requires a calculation based on the UUT and standard specifications.

$$\text{Test Limit}_{\text{IN TOL}} = \text{UUT}_{\text{spec}} - \frac{(\text{Standard}_{\text{spec}}) * 1.6448}{2}$$

where the $\text{Standard}_{\text{spec}}$ is expressed with a 95% confidence level.^[2] If a measurement falls within the $\text{Test Limit}_{\text{IN TOL}}$, there is at least a 95% probability that the measurement made is in-tolerance. As noted earlier, the Fluke calibrator's published specifications are expressed with a 99% or 2.58σ confidence level. Consequently, this formula was used to calculate the standard's uncertainty at a 95% confidence level:

$$\text{Standard}_{\text{spec}} = \frac{2.00 * \text{Published}_{\text{spec}}}{2.58}$$

Table 2 shows the guardbands, or new test limits, for each measurement point as a percentage of the UUT_{spec} .

Measurement Point	TUR	ISO	Deaver
0 mV, 20 mV DC range	3.87	79%	80%
0 mV, 200 mV DC range	3.87	79%	80%
100 mV, 200 mV DC range	3.61	77%	80%
100 V, 200 V DC range	3.91	79%	80%
1000 V, 1000 V DC range	3.91	79%	80%
1 A, 2 A DC slow	2.84	71%	80%
1 A, 2 A DC medium	2.97	72%	80%
1 A, 2 A DC fast	3.87	79%	80%
100 Ω, 200 Ω range, 4-Wire	3.94	79%	80%
100 Ω, 200 Ω range, 2-Wire	3.94	79%	80%

Table 2: Guardbands using the ISO and 1998 Deaver methods.

Interestingly, the new test limits of both methods are very close to each other.

Results

The results of this experiment are shown in Table 3, which shows the number of occurrences a UUT was found to be out-of-tolerance (OOT). (For the purposes of this study, a point not in-tolerance is called out-of-tolerance. This paper does not use the “indeterminate” condition defined in the Agy/Wong paper.) In every case but one, the ISO and Deaver method agreed when a measurement was OOT. In six of the ten cases, guardbanding declared the UUT to be OOT when no guardbanding would have found an in-tolerance condition. In four cases, each method gave the same results.

Measurement Point	ISO OOT	Deaver OOT	No GB OOT
0 mV, 20 mV DC range	2	2	0
0 mV, 200 mV DC range	0	0	0
100 mV, 200 mV DC range	5	5	5
100 V, 200 V DC range	6	6	3
1000 V, 1000 V DC range	5	5	5
1 A, 2 A DC slow	2	0	0
1 A, 2 A DC medium	1	1	0
1 A, 2 A DC fast	1	1	0
100 Ω , 200 Ω range, 4-Wire	0	0	0
100 Ω , 200 Ω range, 2-Wire	3	3	0
Reject (OOT)Totals	25	23	13
Reject Percentages	8.3%	7.6%	4.3%

Table 3: Points considered OOT by the different guardband methods.

Gaining insight into the discrepancies is best done graphically. The 0 mV measurement, shown in Figure 1, indicates that the probability distribution appears normal (i.e. Gaussian) with very little bias. The two guardband methods give the same results, not surprising since one is 79% and the other is 80%. The ISO method declares that a $-0.0029 \mu\text{V}$ reading (which is 97% of UUT_{spec}) has less than a 95% probability the UUT is in-tolerance using the Fluke 5520A. The 1998 Deaver method, with a 80% guardband and a TUR of 3.87:1, states that for a normally distributed UUT, there is a worst case probability of 3% that the true value exceeds the UUT specifications.^[5] Using tables from Deaver’s earlier papers, a 80% guardband also implies that in a population of instruments 7.5% of the calibrations would be declared bad when they were really good.^[3] In a situation like the 0 mV reading, where the population is normally distributed, the “outliers” indicate that the UUT is most likely OOT. An independent check with a higher accuracy source confirmed this.

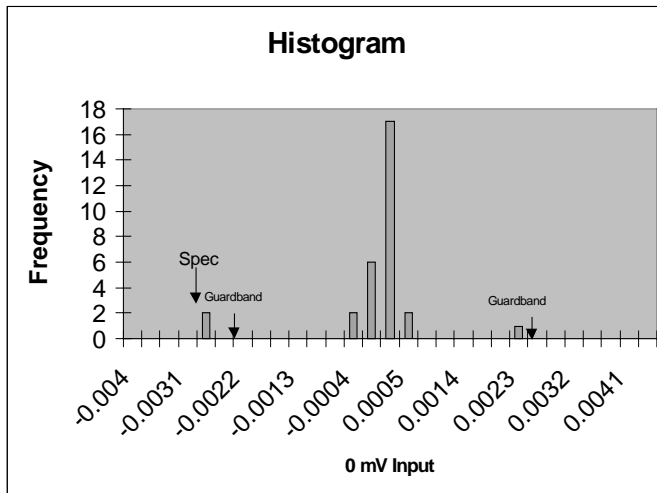
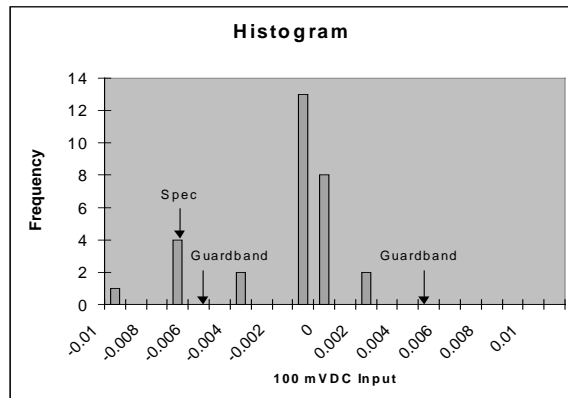
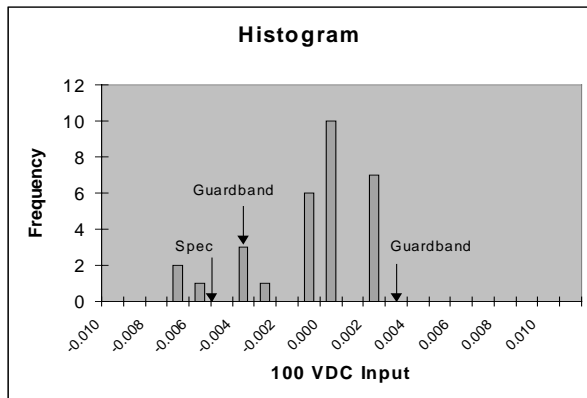


Figure 1.

The 100 VDC and 100 mVDC cases (see figures below) are very similar to 0 mV, with both guardband methods in agreement. The population is normally distributed except for the distinct tail on the left side. The 100 VDC case had 6 cases reported OOT by both guardband methods, but only three when no guardband was applied. An independent check with a more accurate standard was not made at 100 VDC. With an 80% guardband, Deaver would predict a false rejection rate of 7%, or two trials out of the 30. The 100 mVDC case had five OOT declarations, with and without guardbanding.



Another interesting case is the 2 A DC range. This has a low TUR of 2.84:1, giving an ISO guardband of 71%. The Deaver guardband remains at 80%. The calibration trials were skewed in this case by placing the calibrator inside a 0°C chamber. There were two measurements that were 73% of the UUT specification. The ISO guardband declared these occurrences to be OOT, while the Deaver method and no guardbanding declared an in-tolerance condition. Measurements made independently, using much lower measurement uncertainties, indicated the ISO method correctly found the OOT conditions.

In the 2 A DC example with the Deaver guardbanding, what is the calculated consumer risk and producer risk? The chance of declaring a bad calibration good (consumer risk) is .2% and the chance of declaring a good calibration bad

(producer risk) is 9%.^[3] With the 30 trials, there should not have been any observed consumer risk. However, the distribution of this population of measurements approaches a rectangular distribution, not the normal bell curve. (See Figure 4 below.) This adversely affected the calculated vs. observed risk.

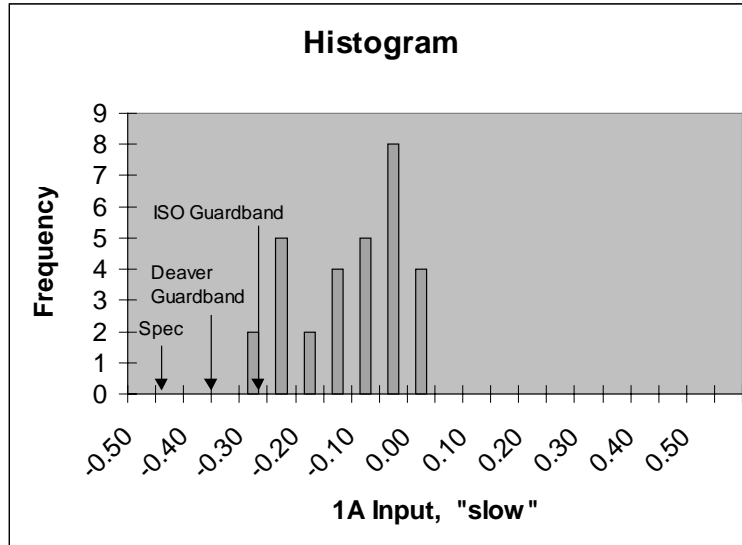


Figure 4.

Guardbanding yields can be difficult to predict. The earlier Deaver papers indicate that an 80% guardband gives a combined rejection rate (which includes actual OOT conditions, false rejects, and false accepts) of 14%. In this study, the total rejection rate for 300 measurements was less than 8%, as shown in Table 3. This is because the typical performance of each measurement is much better than the UUT and standard specifications, pointing out the importance of choosing instruments that have normal distributions with little bias.

Lessons Learned

Is guardbanding effective? Yes, it increases the confidence of making a good measurement. It can be used to quantify risks, just like the older aerospace/military rule of 4:1 TURs. This study has provided several insights:

Lesson 1: Data collection is important to determine the efficacy of guardbanding, especially when the distribution of the UUT is unknown. Where data is available from the metrology community, in published reports or other studies, the probability distribution of UUTs can determine which guardband method to use. Published studies have determined the performance of a population of many instruments (e.g. Fluke 884x series, Keithley 2000, and HP 34401 5½ and 6 ½ digit DMMs).^[6] These popular instruments have little bias and are normally distributed. Hence, Fluke is presently considering using the 1998 Deaver method (80% of specification for TURs less than 4:1) in the Fluke 5520A MET/CAL procedures for these DMMs.

Data collection is simplified with calibration software and computer tools. In this study, results from Fluke's MET/CAL and Microsoft Excel (using the "Add-Ins, Analysis Tool Pak" under the "Tools" menu) were used to determine biases and non-normal distributions. Data collection is not at all inconsistent with establishing a competent laboratory as described in ISO 17025.

Lesson 2: Calibration is about risk management. The choice of guardbanding methods should be dictated by customer requirements. Mission critical applications, like air and space travel, should error on the conservative side. Lacking data, guardbanding may be an evolutionary process. The more conservative ISO method assumes nothing about the population of instruments or measurements, and can be used to quantify the risk of a given measurement. If measurements and instruments are known to be normally distributed, then Deaver's 80% rule may be appropriate.

References

[1] ISO/DIS 17025 – General Requirements for the Competence of Testing and Calibration Laboratories. CS981040033.

[2] Agy, Dave and Wong, Warren; “A Statistical Approach to Calibrate High End Multimeters”, Fluke Corporation, Measurement Science Conference, 1997.

[3] Deaver, Dave; “How to Maintain Your Confidence”, Fluke Corporation, Everett, Washington.

[4] Deaver, Dave; “Managing Calibration Confidence in the Real World”, Fluke Corporation, Everett, Washington.

[5] Deaver, Dave; “Guardbanding and the World of ISO Guide 25. Is There Only One Way?”, Fluke Corporation, NCSL 1998.

[6] Dorchak, Alex; “Extending the Workload Coverage of a Calibrator Using Automated Techniques”, Fluke Corporation, 1997 Measurement Science Conference.