

“Accuracy” Is Two Things, and Calibration Removes Only One

— Bias Does Not Fall with Repetition, Scatter Does Not Fall with Calibration — And What Counts as Bias Is
Decided by What You Repeat — [Paper 255]

Yuuki YAMAGISHI (山岸 勇輝)
Yamagishi Complexity Torus Lab, Japan
August 2026

— *From thinking to syncing.* —

Abstract

The single phrase “this instrument is accurate” names two different quantities. One is the offset from the true value; the other is the scatter. This paper asks whether the two can be reduced by the same operation—the answer is no. Repeating reduces only one of them; calibrating reduces only the other. And which one is the bias and which one is the scatter turned out not to be a property of the instrument, but a consequence of what one chooses to repeat. No new mathematical theorem and no new law is claimed. First, the tools were built here and checked against known values. Evaluating $\text{MSE} = b^2 + \sigma^2$ with 10^6 random draws gives **0.042462**, agreeing with the exact 0.042500 to four decimal places (Section 2). Second, this is the core of the paper. Repetition reduces only one of the two. Averaging 10^6 readings of an instrument with bias 0.05 and scatter 0.20 leaves the error at **0.050000**—it never falls below 0.05 (Section 3). Third, two instruments with the same error part company under averaging. At $n = 1$ both read 0.206155; at $n = 10^4$ they read **0.050040** and **0.200001**—a factor of 3.9968 (Section 4). Fourth, calibration removes the bias alone. The same calibration is worth a factor of 1.0308 at $n = 1$ and a factor of **25.02** at $n = 10^4$ —its value differs by a factor of 24 (Section 5). Fifth, calibration has a price. If the standard is read only m times, calibration imports an error of its own. For $m \leq 16$ it was more accurate not to calibrate at all (Section 5). Sixth, what counts as bias is decided by what one repeats. For the same 100 measurements, one laboratory measuring 100 times gives 0.350571, while 100 laboratories measuring once each give **0.040311**—a factor of **8.70** (Section 7).

Scope of this paper (scope note): No new mathematical theorem and no new law is claimed—that the mean squared error splits into a squared bias and a variance, that the variance of a sample mean is σ^2/n , that the error variance of uniform quantization is $q^2/12$, and that variance separates into levels, are all standard. No theory of measurement is built—the tools used are one decomposition, one sum of two variances, and random numbers. Estimation theory is not entered—unbiasedness, efficiency and the Cramer-Rao fence are treated in Papers 133 and 171. What this paper looks at is an instrument, not an estimator, and the bias here belongs to the instrument, not to the sample. Inverse problems are not entered—Papers 106 and 159 treat them. This paper looks only in the forward direction, at settings where the true value is known. No standard is interpreted clause by clause—the standards are cited only as evidence that the vocabulary itself splits in two, and no question of conformity is discussed. No single Japanese rendering is settled on—the mapping between the Japanese words for accuracy, trueness and precision differs between sources. This paper does not try to repair that disagreement, and writes b and σ instead. Normality is not assumed—the decomposition in Section 2 is an identity and does not depend on the shape of the distribution; the random numbers only confirm it. The numbers involve sampling and discretization—the Monte Carlo uses 10^6 draws, the dither test uses 200 trials, the resolution is $q = 0.10$, and the random seed is fixed. Relation to earlier papers: Paper 133 showed that the Cramer-Rao fence is unbiasedness—this paper follows, in numbers, what lies outside that fence: the case where a bias remains. Paper 171 set unbiasedness and efficiency side by side as separate properties—this paper writes the same two for an instrument rather than an estimator. Paper 106 treated the Wiener filter as the minimizer of the mean squared error—this paper looks at the same MSE from the side of decomposition rather than minimization. Paper 112 counted resolution as one of six distinct roots—Section 6 here observes that this same resolution behaves exactly like a bias in one respect: it does not fall with repetition. Paper 190 measured “rare” on a logarithmic scale—this paper likewise reports the unreachable limit not as a ratio but as the n at which the crossover happens. Paper 232 counted four different claims written with the same equals sign—this paper counts two or more different quantities written with the same word “accuracy”. Paper 186 showed that a proxy stops being a proxy once it becomes the target—Section 7 here observes that changing the unit of repetition exchanges bias and scatter. What is added is building the decomposition of the MSE here and checking it against random numbers, putting a concrete number, $m = 16.67$, on the break-even point of calibration, observing that the value of one and the same calibration ranges from a factor of 1.03 to a factor of 25.02 depending on n , computing that the same 100 measurements differ by a factor of 8.70 according to how they are divided, and stating that “bias or scatter” is settled not by the instrument but by the unit of repetition.

1. Notation and Placement

Setting: a quantity with true value θ is measured by a single instrument. Write a reading as

$$X = \theta + b + \varepsilon, \quad E[\varepsilon] = 0, \quad \text{sd}(\varepsilon) = \sigma.$$

Here b is the part that departs in the same direction every time, and ε is the part that changes every time.

Question: can the two be reduced by the same operation?

The answer is no.Repeating reduces σ alone; calibrating reduces b alone.There are two knobs, and each is connected to something different.

On vocabulary: in English, accuracy is divided into trueness and precision. The Japanese renderings are assigned differently by different sources—one word may be the umbrella term in one text and the name of the scatter alone in another. This paper settles on no rendering and writes b and σ .The disagreement among the renderings is itself evidence that one word is naming two things.

2. Building the Tools Here and Checking Them against Known Values

Check 2.1: take $b = 0.05$ and $\sigma = 0.20$ and evaluate $E[(X - \theta)^2]$ with random numbers. The value to compare against is $b^2 + \sigma^2 = 0.042500$.

Draws	Value from random numbers	Difference from exact
10^5	0.042206	−0.000294
10^6	0.042462	−0.000038

Observation 2.2

At 10^6 draws the difference is 0.000038. The tools used below are calibrated by this.

Observation 2.3: the decomposition is an identity.

$$E[(X - \theta)^2] = (E[X] - \theta)^2 + \text{Var}(X) = b^2 + \sigma^2.$$

No property of the distribution has been used—normality is not assumed.This is the same figure as Paper 171, which wrote that the fence around least squares is linearity rather than normality: here too, what does the work is not the distribution but only the definitions of expectation and variance.

3. The Core: Repetition Reduces Only One of the Two

Setting: measure n times with the same instrument and average. **The bias stays at b under averaging**, while **the scatter shrinks to $\sigma/\sqrt{n}$** .

Check 3.1: computed with $b = 0.05$ and $\sigma = 0.20$. The last column is $\sqrt{b^2 + \sigma^2/n}$.

n	Bias	Scatter	Error
1	0.0500	0.200000	0.206155
4	0.0500	0.100000	0.111803
16	0.0500	0.050000	0.070711
100	0.0500	0.020000	0.053852
10^4	0.0500	0.002000	0.050040
10^6	0.0500	0.000200	0.050000

Observation 3.2

The left column does not move. Multiplying n by 10^6 leaves it at 0.05—repetition does not touch this column.

Check 3.3: solve backwards for the number of repetitions needed to reach a target error, $n = \sigma^2/(\text{target}^2 - b^2)$.

Target error	Repetitions needed
0.070711	16
0.055	77
0.051	397
0.0501	3997
0.05001	39997
0.05	unreachable, however many are taken

Observation 3.4

The closer the target is to 0.05, the more the required count grows by orders of magnitude. Each further decimal place costs roughly a factor of 10 in repetitions, and at 0.05 itself the denominator becomes 0—this is not “slow” but “unreachable”.

Observation 3.5

The crossover is at $n^* = \sigma^2/b^2$. Here $0.04/0.0025 = 16$ —for $n < 16$ the scatter dominates, and for $n > 16$ the bias does. Which of the two should be repaired first is exchanged by how many measurements one intends to take.

Observation 3.6

Do not report a ratio. “How many times worse is the bias than the scatter” moves with n and is therefore not a single number—what should be reported is the n at which they cross (the same practice as Paper 190).

4. Two Instruments with the Same Error Part Company under Averaging

Setting: take two instruments. A has $b = 0.05$, $\sigma = 0.20$; B has $b = 0.20$, $\sigma = 0.05$. At $n = 1$ the errors of the two are identical.

Check 4.1:

n	Error of A	Error of B	Ratio
1	0.206155	0.206155	1.0000
10	0.080623	0.200624	2.4884
100	0.053852	0.200062	3.7151
10^4	0.050040	0.200001	3.9968
∞	0.050000	0.200000	4.0000

Observation 4.2

Even when the single-reading error quoted on the data sheet is the same, a difference in use makes them differ by a factor of 4. A data sheet that writes “accuracy” as one number cannot express this difference.

Check 4.3: for four instruments, count how much averaging 10^4 readings is worth.

Instrument	b	σ	Error at $n = 1$	Value of averaging
Small bias, small scatter	0.01	0.02	0.022361	2.2356
Small bias, large scatter	0.01	0.50	0.500100	44.7303
Large bias, small scatter	0.50	0.02	0.500400	1.0008
Large bias, large scatter	0.50	0.50	0.707107	1.4141

Observation 4.4

The second and third instruments have single-reading errors of 0.500100 and 0.500400, which are all but the same. Yet the same average over 10^4 readings is worth a factor of 44.73 to one and only 1.0008 to the other—the return on the same operation differs by a factor of 44.

Observation 4.5

The separator—what has to be changed for these to become different things. Not the single-reading error, but how that error is composed. Hand over one number without the composition, and what to do next cannot be decided.

5. Which of the Two Calibration Removes, and What Calibration Costs

Setting: measure a standard whose true value is known, and subtract the departure. **This is calibration.**

Check 5.1: the ideal case, in which the bias is known exactly. Then $b = 0.05 \rightarrow 0$, while σ stays at 0.20 and does not move.

n	Before calibration	After calibration	Value
1	0.206155	0.200000	1.0308
10^4	0.050040	0.002000	25.02

Observation 5.2

The value of the same calibration ranges from a factor of 1.03 to a factor of 25.02 according to n . If only one reading is taken, calibration is worth almost nothing; if 10^4 readings are taken, calibration is almost the whole of the gain—the value of calibration is set not by the instrument but by how many measurements are still to come.

Check 5.3: a real calibration estimates the departure from m readings of the standard. The estimate imports an error of its own: $\sigma^2/m + u^2$, where $u = 0.01$ is the uncertainty of the standard itself. Compare this with $b^2 = 0.0025$, the case of not calibrating.

m	Error imported	Against b^2	Floor after calibration
4	0.01010000	worse	0.100499
9	0.00454444	worse	0.067412
16	0.00260000	worse	0.050990
17	0.00245294	better	0.049527
25	0.00170000	better	0.041231
100	0.00050000	better	0.022361
10^4	0.00010400	better	0.010198

Observation 5.4

The break-even point is $m = \sigma^2/(b^2 - u^2) = 16.67$. If the standard is read 16 times or fewer, it was more accurate not to calibrate—calibration is not an operation that improves things unconditionally.

Observation 5.5

However large m is made, the floor stops at $u = 0.01$. The uncertainty of the standard itself cannot be moved by repetitions on this side—calibration does not remove the bias; it replaces one's own bias with the bias of the standard.

Observation 5.6

The two knobs do not meet. Repetition is connected only to σ , and calibration only to b . However far one is turned, the floor set by the other does not move.

6. Bias Is Not the Only Thing That Does Not Fall with Repetition

Setting: a display with graduation $q = 0.10$ reads a true value of **20.037**. The reading is rounded to the nearest graduation.

Check 6.1: read the same value repeatedly on an ideal instrument with no scatter.

Readings	Averaged reading	Error
1	20.000000	−0.037000
100	20.000000	−0.037000
10^6	20.000000	−0.037000

Observation 6.2

It never falls. With zero scatter, averaging only returns the same value again—here the error due to resolution behaves entirely as a bias.

Check 6.3: add scatter of standard deviation $\sigma = 0.20$ before the same display reads (that σ is twice the graduation q). The error is measured over **200** trials.

Readings	Error	$\sigma/\sqrt{n}$
1	0.199772	0.200000
100	0.020192	0.020000
10^4	0.002142	0.002000

Observation 6.4

Now it falls as $\sqrt{n}$. The averaged answer comes closer to the true value when scatter is added—the same graduation on the same display became a bias in one case and a scatter in the other.

Observation 6.5

What separated them was not the instrument. It was whether the scatter is larger than the graduation—at $\sigma = 0$ it behaves as a bias, and at $\sigma \gg q$ as a scatter. Paper 112 counted resolution as one of six distinct roots—what this section adds is the single point that the same resolution moves to the bias side or to the scatter side.

7. What Counts as Bias Is Decided by What You Repeat

Setting: several laboratories measure the same sample. The scatter now has two storeys: the scatter on repetition within one laboratory, σ_r , and the standing departure of each laboratory, σ_L . Write the combination as σ_R .

Check 7.1: take $\sigma_r = 0.20$ and $\sigma_L = 0.35$.

Quantity	Value
Scatter within one laboratory σ_r	0.200000
Departure of each laboratory σ_L	0.350000
Combination $\sigma_R = \sqrt{\sigma_L^2 + \sigma_r^2}$	0.403113
Ratio σ_R/σ_r	2.015564

Check 7.2: increase the number of repetitions within one laboratory.

n	Error from the true value
1	0.403113
4	0.364005
16	0.353553
100	0.350571
10^4	0.350006

Observation 7.3

It stops at 0.35. So long as one stands inside a single laboratory, that laboratory's departure is a bias—it does not fall with repetition. The table has the same shape as the one in Section 3.

Check 7.4: take the same 100 measurements and change only how they are divided between laboratories and repetitions. Every row totals 100.

Division	Total	Error
1 laboratory $\times$ 100 readings	100	0.350571
2 laboratories $\times$ 50 readings	100	0.248294
5 laboratories $\times$ 20 readings	100	0.157797
10 laboratories $\times$ 10 readings	100	0.112472
25 laboratories $\times$ 4 readings	100	0.072801
100 laboratories $\times$ 1 reading	100	0.040311

Observation 7.5

The same 100 measurements differ by a factor of **8.70**. The number of measurements is the same and the effort spent is the same—what differs is only what was repeated.

Observation 7.6

This is the core of the paper. The laboratory departure σ_L is a bias when one stands inside a single laboratory (it does not fall with repetition) and a scatter when one stands across laboratories (it falls with the number of laboratories). One and the same quantity becomes either. “Bias or scatter” is not a property of the instrument but a question of where the unit of repetition has been placed.

Observation 7.7

A point to state honestly—if every laboratory calibrates against the same standard, σ_L does shrink. But that means the departures between laboratories were a collection of biases. What remains after the shrinking is the true scatter at that level—calibration is also the experiment that separates the two.

8. Roll Call and Closing

Quantity	Falls with repetition?	Falls with calibration?
Bias b	no	yes
Scatter within one laboratory σ_r	yes	no
Departure of each laboratory σ_L	depends on the unit of repetition	shrinks against a common standard
Resolution q	depends on the size of the scatter	no

Observation 8.1

Of the four rows, only the top two admit an answer written as “yes” or “no”. The lower two become either, according to the conditions.

1. The tools were calibrated (Check 2.1): a difference of 0.000038 from the exact value at 10^6 draws.
2. The decomposition is an identity (Observation 2.3): no property of the distribution is used.
3. Repetition reduces only one (Check 3.1): averaging 10^6 readings still stops at 0.050000.
4. The crossover is at $n^* = 16$ (Observation 3.5): which to repair first is exchanged by n .

5. Two instruments with the same error part company (Check 4.1): a factor of 3.9968 at $n = 10^4$.
6. The return on averaging differs by a factor of 44 (Check 4.3): 44.7303 against 1.0008.
7. The value of calibration is set by n (Check 5.1): 1.0308 against 25.02.
8. The break-even point of calibration is $m = 16.67$ (Observation 5.4): at 16 readings or fewer it is more accurate not to calibrate.
9. Resolution moves to either side (Checks 6.1 and 6.3): it stays at -0.037000 , or it falls as $\sqrt{n}$.
10. The same 100 measurements differ by a factor of 8.70 (Check 7.4): 0.350571 against 0.040311.

Closing: “this instrument is accurate” was not saying one thing. It named the departure from the true value and the scatter—the first does not fall with repetition, and the second does not fall with calibration. The two knobs do not meet, and however far one is turned the floor set by the other does not move. And which of them is the bias and which the scatter turned out not to belong to the instrument. Add scatter and the resolution turns into scatter; move the unit of repetition to the laboratory and the laboratory's departure turns into scatter. One thing separates them—writing down what it is that is being repeated. Write it down, and which knob to turn next is settled. Do not write it down, and one cannot see that 100 measurements have been spent at a loss of a factor of 8.70.

References

1. ISO 5725-1:1994. *Accuracy (trueness and precision) of measurement methods and results — Part 1: General principles and definitions*. Geneva: International Organization for Standardization.
2. ISO 5725-2:1994. *Accuracy (trueness and precision) of measurement methods and results — Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method*. Geneva: International Organization for Standardization.
3. JCGM 200:2012. *International Vocabulary of Metrology — Basic and General Concepts and Associated Terms (VIM)*, 3rd edition. Sevres: Joint Committee for Guides in Metrology.
4. JCGM 100:2008. *Evaluation of Measurement Data — Guide to the Expression of Uncertainty in Measurement (GUM)*. Sevres: Joint Committee for Guides in Metrology.
5. Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*, 2nd edition. New York: Springer.
6. Taylor, J. R. (1997). *An Introduction to Error Analysis*, 2nd edition. Sausalito: University Science Books.
7. Widrow, B. and Kollar, I. (2008). *Quantization Noise*. Cambridge: Cambridge University Press.
8. Yamagishi, Y. (2026). *Extracting Signal from Noise* [Paper 106]. Zenodo. 10.5281/zenodo.21836984
9. Yamagishi, Y. (2026). *The Allotted Region and the Freedom Within* [Paper 112]. Zenodo. 10.5281/zenodo.21962834
10. Yamagishi, Y. (2026). *The Fisher Metric Is Unique Only on the Classical Side* [Paper 133]. Zenodo. 10.5281/zenodo.22080550
11. Yamagishi, Y. (2026). *“Ill Posed” Is Not One Word* [Paper 159]. Zenodo. 10.5281/zenodo.22092229

12. Yamagishi, Y. (2026). *The Fence around Least Squares Is Linearity, Not Normality* [Paper 171]. Zenodo. 10.5281/zenodo.22094287
13. Yamagishi, Y. (2026). *When a Proxy Becomes the Target It Stops Being a Proxy* [Paper 186]. Zenodo. 10.5281/zenodo.22105530
14. Yamagishi, Y. (2026). *“Rare” Is Measured on a Logarithmic Scale* [Paper 190]. Zenodo. 10.5281/zenodo.22106182
15. Yamagishi, Y. (2026). *Learn from a Book with the Proofs Removed and the Equals Sign Loses Its Type* [Paper 232]. Zenodo. 10.5281/zenodo.22127528

— *From thinking to syncing.* —

Yuuki YAMAGISHI (山岸 勇輝)

© 2026 Yamagishi Complexity Torus Lab, Japan · Licensed under Creative Commons Attribution 4.0 International (CC BY 4.0)