How should e function in the Scales Code, and why it doesn't?
Ross Andersen, (NY retired) December 22, 2023

I submitted a proposal in 2019 that became the Block 2 items and eventually the proposal of the Verification Scale Division e Task Group. I can now see the crux of the issue is not that clearly laid out in either proposal. Namely, what is the verification scale division e, and how does it function? How is it different from the scale division d? In response to comments from a colleague, I will try to answer those questions, by looking at how e is intended to function in R76, the original source of the Scales Code.

> **(R76) T.3.2.2 Actual scale interval, *d* –** Value, expressed in units of mass of: the difference between the values corresponding to two consecutive scale marks, for analog indication; or the difference between two consecutive indicated values, for digital indication.

I begin my analysis with the actual scale interval d, the smallest increment of indication. Take note of the term "scale." I suggest it is not referring to a weighing instrument, since the instrument is not divided. Rather it refers to the "scale of measurement." All indicating instruments use a scale of measurement with that "scale" being a series of analog graduations or digital increments with associated quantity values. To avoid this confusion, R76 uses the term "weighing instrument," not "scale" and not "weighing device." I will follow the R76 convention of calling them weighing instruments, or just instruments. Also take note of the R76 term "interval" which is parallel to the HB44 term "division." These terms both refer to an increment of a measurement scale.

A critical principle underlying the increments of the instrument measurement scale is that they are expected to be uniform in quantity. Consider five consecutive intervals for an instrument where d is rated as 1 g but the weight represented is only 0.8 g instead. This results in a change in the error of 1 g over a 5 d interval. If we cannot accept that scale intervals are uniform, it becomes necessary to test an instrument at many, many points to ensure compliance. R76 clearly expects uniformity in the scale intervals and usually employs only five or fewer test loads to verify conformance.

Instrument 1	Instrument 2	Instrument 3	Instrument 4
cap 20,000 g d = 0.1 g round half up/half down	cap 20,000 g d = 1 g round half up/half down	cap 20,000 d = 20 g round up	cap 20,000 g d = undefined rounding decided by user

These four weighing instruments provide excellent examples to help discuss weighing instruments. The capacities are all 20,000 g, but the scale divisions range from 0.1 g to 20 g, or undefined. Also notice the rounding. Instruments 1 and 2 have normal, half up/half down rounding. Instrument 3 is what we call a weight classifier, rounding up. Instrument 4 is a balance, where the manufacturer makes no claim of the size of d. Resolution, the size of d, is based on the increment of counterbalance weights chosen by the user. With a balance, the user also decides the rounding method. With a tipping point, the user can either round up or down. With balancing, the rounding is typically half up/half down.

Looking at the capacity and division size (resolution) ratings, which of those four instruments is more accurate? A common response is that 1 is more accurate than 2, and 2 more accurate than 3, based on the smaller value of d. This is a misunderstanding, since measurement science tells us that resolution (the size of d) is unrelated to accuracy. More important, based on the HB44 definition, there is insufficient information provided to answer the question.

> (HB44 App D) accurate. **–** A piece of equipment is "accurate" when its performance or value – that is, its indications, its deliveries, its recorded representations, or its capacity or actual value, etc., as determined by tests made with suitable standards - conforms to the standard within the applicable tolerances and other performance requirements. Equipment that fails so to conform is "inaccurate."

Assessing accuracy compliance requires knowledge of the applicable limits of error (tolerances). The resolution (d) is unrelated and unconnected to these limits. Also note that HB44 "accuracy" applies to a group of instruments with the same accuracy parameters. Within this group, all instruments performing within the same applied tolerances are equally accurate.

With this definition, one instrument in the group cannot be more accurate than another, just because it has smaller errors in the tests conducted. Conformance of a member of the group is all or nothing, not a degree. Only another group, with different parameters, can be more or less accurate. Hence class I is more accurate than class II.

Within HB44, the tolerance for most instruments is based on application. If the instrument is used in a different application, the tolerances change accordingly. The HB44 Scales Code and section 2.24. the AWS Code do not follow that pattern. Like R76, in these Codes the manufacturer declares the accuracy (by classification markings) and the tolerances do not change with different application. The verification scale interval (or division) e is one key element of this declaration.

> (R76) T.3.2.3 Verification scale interval, e – value, expressed in units of mass, used for the classification and verification of an instrument.

Notice first that in this definition e there is no stated nor implied connection between e and the indications of a weighing instrument (contrast with the definition of d). We will find e is not an indication. Consider the two uses of e.

Classification happens early in the design phase. The manufacturer determines the class before specifying the appropriate measurement principle and the components that will be used to construct the instrument. The principles are stated in R76 2.2 and I've included the definition of n for reference.

> (R76) 2.2 Principles of the metrological requirements **….**Instruments are classified according to: the verification scale interval, representing absolute accuracy; and the number of verification scale intervals, representing relative accuracy. …

> (R76) T.3.2.5 Number of verification scale intervals, *n* - Quotient of the maximum capacity and the verification scale interval: $n = \text{Max} / e$

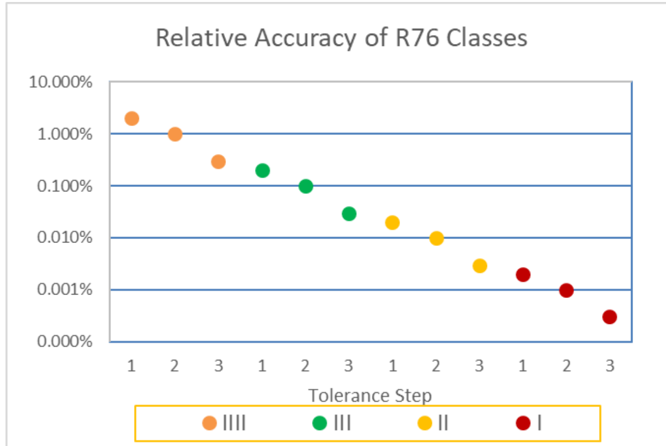
Table 3 specifies the classification parameters. The manufacturer chooses the class and the values for capacity and e. Dividing capacity by e you get n, and n must fit within the range permitted for the class in columns 3 and 4. The instrument must also be capable of meeting error limits in Table 6. I will address column 5 and the minimum load later.

(R76) Table 3

Accuracy class	**Verification scale interval, *e***	**Number of verification scale intervals, *n* = Max/*e***		**Minimum capacity, Min (Lower limit)**
		minimum	maximum	
Special (I)	$0.001\text{ g} \le e$*	50 000**	–	100 e
High (II)	$0.001\text{ g} \le e \le 0.05\text{ g}$ $0.1\text{ g} \le e$	100 5 000	100 000 100 000	20 e 50 e
Medium (III)	$0.1\text{ g} \le e \le 2\text{ g}$ $5\text{ g} \le e$	100 500	10 000 10 000	20 e 20 e
Ordinary (IIII)	$5\text{ g} \le e$	100	1 000	10 e

Notice the two kinds of accuracy in 2.2, absolute accuracy and relative accuracy. These are important distinctions. What permits an instrument to fit in a class is the relative accuracy, not the absolute accuracy. Absolute accuracy is expressed by the manufacturer in the declared value of e. The e values in the four R76 classes overlap to large extents. For example, any e of 5 g or larger can be found in all four classes, and values down to 1 mg can be found in both Classes I and II.

In contrast, relative accuracy is expressed in terms of two values, e and n, practically manifest as tolerance and test load. The relative accuracies of the R76 classes, as shown in the figure at right, do not overlap. You can compute relative accuracy at the maximum test load for each tolerance step based on R76 Table 6. Class IIII ranges from 2% accurate in the first step (1e tolerance for 50 e load), to 1% in the second step (2 e for 200 e), to 0.3% in the third step (3 e for 1,000 e). This pattern repeats but the relative accuracy increases by a factor of 10 as you move to each higher accuracy class.



Verification is the process of ensuring the instrument performs within the accuracy claims. Verification happens after the instrument is produced for type evaluation and after being placed in service for initial and subsequent verification. The verification scale interval e is the tool used in choosing test loads and expressing tolerances as indicated in Table 6. Metrologically, you need a reference (test load) that is external to the device to measure accuracy.

(R76) Table 6

Maximum permissible errors on initial verification	**For loads, *m*, expressed in verification scale intervals, *e***			
	Class I	**Class II**	**Class III**	**Class IIII**
± 0.5 *e*	$0 \le m \le 50\ 000$	$0 \le m \le 5\ 000$	$0 \le m \le 500$	$0 \le m \le 50$
± 1.0 *e*	$50\ 000 < m \le 200\ 000$	$5\ 000 < m \le 20\ 000$	$500 < m \le 2\ 000$	$50 < m \le 200$
± 1.5 *e*	$200\ 000 < m$	$20\ 000 < m \le 100\ 000$	$2\ 000 < m \le 10\ 000$	$200 < m \le 1\ 000$

Returning to our four instruments we can add class and e and re-ask, which of these four instruments is more accurate?

Instrument 1	Instrument 2	Instrument 3	Instrument 4
cap 20,000 g d = 0.1 g class II e = 1 g	cap 20,000 g d = 1 g class II e = 1 g	cap 20,000 d = 20 g class II e = 1 g	cap 20,000 g d = undefined class II e = 1 g

Based on the new information we find all four instruments have the same accuracy parameters and thus <u>the same accuracy</u>. This exercise was to help distinguish between resolution (measured in d) and accuracy (measured in e). This confirms that resolution is disconnected from accuracy. This is why we need a separate value (e). This is clear in R76 from Table 2 and a note to Table 3 where the four kinds of instruments are specifically mentioned. I added numbers referring to the four examples used above. This information (also lost in translation) is summarized below.

Kind of Instrument	e vs d	Description, resolution and rounding
1. High Resolution	e > d	with auxiliary indication, resolution d with half up/half down rounding
2. Normal Resolution	e = d	without auxiliary indication, resolution d with half up/half down rounding
3. Low Resolution	e ≤ d	weight classifier, resolution d with mostly up rounding
4. Undefined Resolution	e no d	balance – resolution and rounding at discretion of user

In the first three examples the resolution d ranges two-hundredfold, from 0.1 g to 20 g. With the balance there is no defined resolution, but it is defined by the user in how the instrument is operated. Yet the relative accuracy of all four example instruments is the same, 2 e per 20,000 e, 2 parts in 20,000, 2 g in 20,000 g, or 0.01%.

This brings us to “error,” the deviation from the true value.

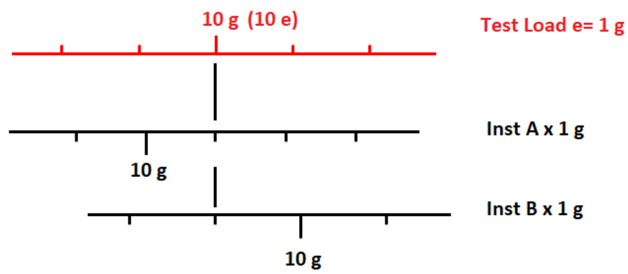
> (R76) T.5.5.1 Error (of indication) **-** Indication of an instrument minus the (conventional) true value of the corresponding mass.

In the verification we are comparing two different things, indication (derived from d) and test load (expressed in e). A key point here is that d and e can never be interchanged in this expression of error. It is also critical to see that it is the indication that is verified, not the value of d nor the value of e. It is best to think of e as a tool used in verification, see definition. Consider a test of instrument 2 at 5,000 g (5,000 e) that produces an indication of precisely 5,001 g (5,001 d). The error is clearly 1 g, but is that error 1 e or 1 d?

Error 1 g (1 d or e?) = indication 5,001 g (5,001 d) – test load 5,000 g (5,000 e)

One way to look at it is to consider the indication and the test load as equivalents, i.e., 5,001 d indication = 5,000 e test load. With this formula you can define d in terms of e (d = 0.9998 e) or define e in terms of d (e = 1.0002 d). It is d defined in terms of e that is meaningful since we know what e is in the certified test weights. We don't know what d is because it is specific to indications of this particular instrument.

A good way to visualize this is using the graphic at right. We see two weighing instruments A and B with applied load in red of 10 g, or 10 e. For instrument A the indication is 11 g (11 d) and for instrument B it is 9 g (9 d).



Next, think of the correction as the amount of change required to bring the error in the indication to zero. If we correct the 11 d indication of instrument A by 1 d (moving the 11 d indication up to where 12 d is), we find the error is not zero as we under-corrected. The d's for instrument A are smaller than the e's (or what the d should be) by about 10%, in this case 11 d = 10 e, or d = 0.909090 e.

However, if we were to correct the 11 d indication by 1 e, the 11 d indication will align precisely with 11 e load and the error becomes zero. Similarly for instrument B, a correction of 1 d downward overcorrects because the d's are larger than the e's (d = 1.1 e). If we correct the 9 d indication by 1 e, the 9 d indication aligns with 9 e and the error becomes zero. Notice in making a correction, all of the divisions d are changed proportionally (remember d's are uniform).

Thus, we find the e's in R76 are used in verification to define the test loads, the errors, and the tolerances. I reiterate these three things are separate from, and independent of, the instrument indications in d. That should have been very evident from the test where the test load of 5,000 e produced an indication of 5,001 d. Under HB44, where 5,000 d test load = 5,001 d indication, we have confusion since this requires there be two kinds of d's.

Always showing errors in whole e increments may be a disservice, as errors occur in whole numbers of e only between 10% and 20% of the time, depending on error resolution. A test load of 5,000 g on the scale may result in an indication of 5,000.6 g, an error of +0.6 g or 0.6 e. R76 clearly states that errors should be resolved to 0.2 e or finer, yet HB44 is silent. The Task Group opted to ignore this subject. When testing most instruments under HB44 the error is resolved to at least 1/5 of the maintenance tolerance. Yet with the Scales Code the tests often resolve only to the maintenance tolerance. Think of that as if you used a prover with a 6 inch neck diameter and 6 cu in graduations to test a fuel dispenser.

You can resolve errors to finer than 1 e in multiple ways per R76. For analog, there are three options. (1) You can interpolate. (2) You can use error weights to alter the test load to bring the index in coincidence with a graduation. For example, using the example in the previous paragraph, we could add 0.4 g (0.4 e) to result in an indication of 5,001.0 g. Thus, with a test load of 5,000.4 g the error is still +0.006 g (indication 5,001 g – test load 5,000.4 g). (3) You can also use an auxiliary indication (a means to increase resolution) like a vernier on some class I or II instruments to read between the graduations, akin to interpolation.

For digital, there are also three options. (1) You can use error weights with the break point between increments replacing the analog graduation as the reference point. (2) You can use auxiliary indications where d is 0.2 e or smaller on some class I or II instruments. (3) You can use an extended displaying means to temporarily increase the resolution during tests. This third option has been available under R76 for many years. HB44 failed to keep pace.

> (R76) 2.2 Principles of the metrological requirements … The maximum permissible errors are in the order of magnitude of the verification scale interval. They apply to gross loads and when a tare device is in operation they apply to the net loads. The maximum permissible errors do not apply to calculated net values when a preset tare device is in operation.

One final note about errors and tolerance. The application of tolerance to net values reinforces the uniformity of the scale divisions d. When using a R76 tare device (HB44 semiautomatic tare), the net zero is set to the center of the digital division at the tare load (similar to zero with no load). Because scale divisions must be uniform, the instrument should meet the tolerances starting from any net zero indication after use of a tare device. A preset tare cannot guarantee center zero at the tare load, so R76 does not apply tolerances with preset tare (HB44 keyboard or programmed tare).

Returning to column 5 in (R76) Table 3, why is minimum load part of the accuracy classification? It is explained in one part of the principles in section 2.2.

> (R76) 2.2 Principles of the metrological requirements …….. A minimum capacity (Min) is specified to indicate that use of the instrument below this value is likely to give rise to considerable relative errors.

The key elements in this statement are "use of the instrument" and "considerable relative errors." You must differentiate these from accuracy of the instrument, which is verified in testing. In testing per R76, you reduce rounding in the measurement of error to at least 0.1 e and you focus on relative accuracy, as in 1 e per 5,000 e or 0.02%. In commerce you are dealing with loads that are rarely whole e increments, and thus you must deal with rounding error to resolution (usually d). This is the focus of minimum load.

Minimum quantity is an issue in the use of any indicating instrument. Consider a meter rule with 1 mm graduations that is accurate to 1 mm at 1 m, or 0.1%. I hand you a piece of printer paper and ask you to measure the thickness of the paper (about 0.025 mm). The 1 mm divisions are clearly inadequate. They are not inadequate because they are inaccurate! The 1 mm graduation should be roughly accurate to 0.1 % or within 0.001 mm, which you can't see because of the large uncertainties. Rather, they are inadequate because the potential rounding error to the nearest 1 mm graduation is large relative to the quantity. Remember that rounding of +/- 0.5 d per 1 d is 50 % relative error. This is the "considerable relative errors" mentioned in 2.2., but they only occur in use and not in testing.

In R76, minimum load from Table 3 leads to a marking requirement of Min that goes along with class, Max, and e. The markings of Class, Max and e provide the classification information necessary to verify any instrument and the Min provides the information on the lower usable range of the instrument. It is a specification because R76 does not presume to regulate users in the member countries. This is one instance where R76 could be clearer. It expresses Min in column 5 in terms of accuracy (normally e), when minimum load is a resolution issue (normally d). In R76 there are two important exceptions to column 5. The first is in a footnote to Table 3 where it specifies a 5 e Min for grading instruments (weight classifiers). The other exception is found in a subsection of 3.4 which deals with auxiliary indications.

> **3.4.3 Minimum capacity -** The minimum capacity of the instrument is determined in conformity with the requirements in Table 3. However, in the last column of this Table, the verification scale interval, *e*, is replaced by the actual scale interval, *d*.

Thus, we see that sometimes R76 is defining Min in terms of e and sometimes in d? What is the underlying metrological principle? We can discover it by going back to our four kinds of instruments. The R76 values for Min would be 5 g (50 d) for instrument 1, 50 g (50 e) for instruments 2 and 4, and 5 g (5 e) for instrument 3.

Instrument 1	Instrument 2	Instrument 3	Instrument 4
cap 20,000 g d = 0.1 g class II e = 1 g Min = 5 g	cap 20,000 g d = 1 g class II e = 1 g Min = 50 g	cap 20,000 d = 20 g class II e = 1 g Min = 5 g	cap 20,000 g d = undefined class II e = 1 g Min = 50 g

The principle involved can be stated as: resolution limits how small you can measure effectively, due to potential rounding errors to the scale division d. With instrument 1 the resolution is 0.1 g with the auxiliary indication. The potential rounding error is +/- ½ d or 0.05 g. In contrast, with instrument 2 the resolution is 1 g with potential rounding error of +/- ½ d or 0.5 g. R76 understands that e = d for these instruments so the Min (50 e or 50 d) is the same using e or d and that's ten times more rounding error than instrument 1. With instrument 3, the rounding is mostly up, so Min is arbitrarily chosen at 5 e. I suspect this is to ensure the load is sufficiently above zero. Instrument 4 is unusual since the manufacturer does not define the resolution. The used resolution is at the discretion of the user. In this case, R76 based marking of Min on the required sensitivity of the instrument when balancing is used. This implies normal half up/half down rounding. It also recognizes that a balance sensitivity is not unlimited, and sensitivity is expected to be close to 1 e at light loads. This means a 1 e load will cause a minimum permanent change in the balance indicator of 1 increment.

The reasoning behind setting a minimum value for Min is that rounding the indication to the scale division d for any class will not contribute excessive rounding errors in use. For comparison of relative errors, Min for class IIII is ½ d in 10 d or 5%, for class III it is ½ d in 20 d or 2.5%, for class II it is ½ d in 50 d or 1%, and class I it is ½ d in 100 d or 0.5%. Just remember that these are small percentages of small quantities. I stress that minimum load is not an accuracy (e) issue. The relative error of the instrument combined with uniform scale divisions d will never produce large errors in testing in the lower range of the instrument. Try it for yourself. Use error weights and see if you can detect any appreciable error in the vicinity of Min on a commercial instrument. For example, a class II instrument should be accurate to ~0.02 % in the first tolerance step. When e = d that is about 0.01 e error at 50 e load, much too small to be seen in testing.

The problem with the current Task Group proposal is how it treats instruments with auxiliary indication. These should be recognized in a note to change Table 8. column 3 to d for these instruments as in R76. Consider an analog indication with a vernier. There has always been some misunderstanding about these instruments, revolving around the idea that the auxiliary indications on the vernier were somehow inferior to the primary graduations. R76 and HB44 both have requirements on d such as the uniformity of graduations G-S.5.2.3., the clear space between graduations S.1.3.2., and discrimination T.N.7. The d's with auxiliary indication must comply with all the requirements on d. If you think about it, the auxiliary indications are no less, and no more, accurate than any other division on the measurement scale. If e = 10 d, stipulate a load of 5,000 e produces an indication of 50,006 d. What if the test load is changed to 5,009 e, 4,988 e or 4,999.5 e. With uniform divisions, the errors observed at all of these loads should be the same +0.6 e as at 5,000 e. This applies equally to digital indications when you resolve to 0.2 e or better.

The fix is simple. One possibility is to add a second note to Table 8. for instruments with auxiliary indications to change. column 3 to d just as it is done in R76. The other possibility is to make column 3 in d with notes to change d to e for balances and use 5 e for weight classifiers. The second option was a previous Task Group decision until a later change. The advantage of using d in column 3 is that it emphasizes minimum load is a resolution issue and not an accuracy issue.

I can't stress enough that the verification scale division e is something separate from, and independent of, the scale division d. These two are defined differently, and in R76 they can never be interchanged. The e value is used by the manufacturer in classification and the official in verification (testing). The d is simply the smallest part of the indication of the instrument that is being verified. This is born out further in the way we typically graph instrument error. The X axis is the test load expressed in e. The Y axis is the error, i.e., indication minus test load, and the tolerance which are both expressed in e. Notice carefully that the error is based on the indication, which is the product of multiple d's.

What everyone needs to understand from this analysis is that the Scales Code in HB44 is confusing at best and wrong at worst, because it lost the clear distinction between e and d. It does this since several key paragraphs were mistranslated, either sending mixed messages or wrong messages. The Task Group recommendations combined with some minor retraining will remedy this situation.