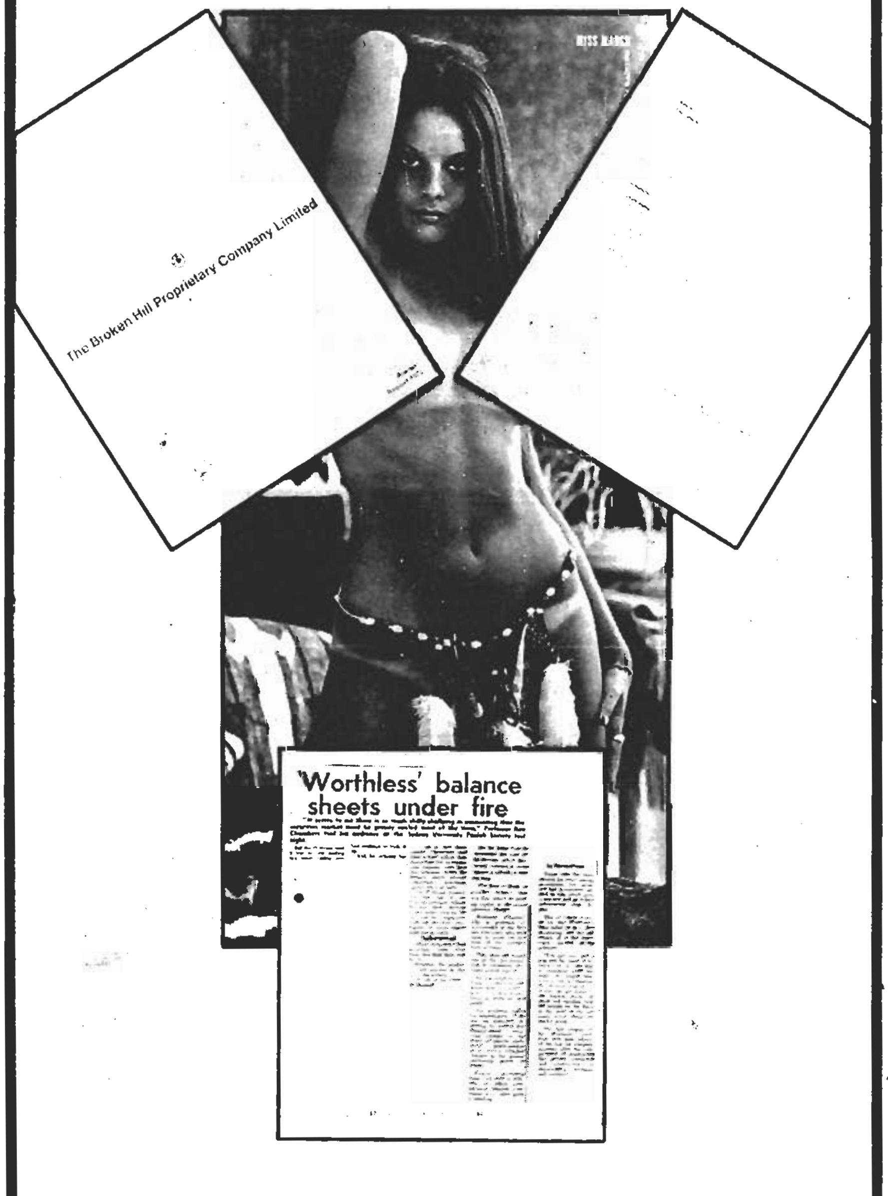
Measurements

Misunderstandings, misuse, mistakes, mirth and misrepresentation

by Dr. Peter Sydenham



The company balance sheet – what the numbers reveal is interesting but what they hide is vital. (Courtesy John Staunton)

MEASUREMENTS are useful as tools of control of the routine, or as a basis for gaining new knowledge.

Once a particular kind of measurement has been established as necessary, two things can happen. It either remains as is, being adequate for the daily need, or alternatively, constant effort is made to improve it in order to gain the benefits that might accrue from a better determination.

WHAT IS A MEASUREMENT?

To some people measuring implies attaching sophisticated black boxes to a system in order to obtain data about it. Others see the opposite – the use of simple tools such as a ruler to put data to an object. Both are right in a narrow sense. A first basic rule is that *measurement is the comparison of an unknown magnitude of a quantity with an agreed standard declared as the unit, the measure coming forth as the difference expressed in numerical form.*

The most basic number is a binary kind having just two states. The crudest measurement we can thus make is one that provides a yes or no, smaller or larger, up or down, go or no-go, true or false types of answer. Enormous effort can be expended to obtain (or try to obtain) such an answer in many cases. Indeed, it is this kind of measurement that is often the hardest to make. Politicians would give the earth to be able accurately to predict the outcome of an election, social-scientists would be enthralled at the prospects of certainty of success of implemented crime-control measures, geographers seek to know which factors affect what.

As the understanding of a subject is improved – by the use of simple tests giving yes-no type answers – it becomes possible to deploy more and more advanced techniques of hardware and software. The difference between the standard and the unknown (called the measurand) becomes expressible in continuous rather than two-state number terms. For example, the

Measurements

battery in the car will not turn the starter motor because its voltage is *too low* (the two-state situation has been extended by a superlative). Measurement using a voltmeter enables us to say that the battery output is only 6.6 V instead of 12.0 V (the designer's standard requirement now expressed in numbers on a continuous scale).

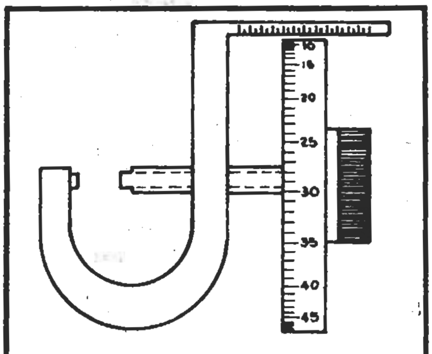


Fig. 2. Adding a huge drum to the dial of a screw micrometer increases its resolution but not necessarily the precision and accuracy of measurements made.

BASIC MEASUREMENT TERMS

Even though we have not yet talked in terms of sophisticated measurements or measuring equipment, the above explanation begins to reveal the need for some definitions of certain facets arising in the intercomparison process. (We need more standards to define standards!)

The misuse and abuse of basic terminology is rife. A first group of terms relates to the description of results provided by the process.

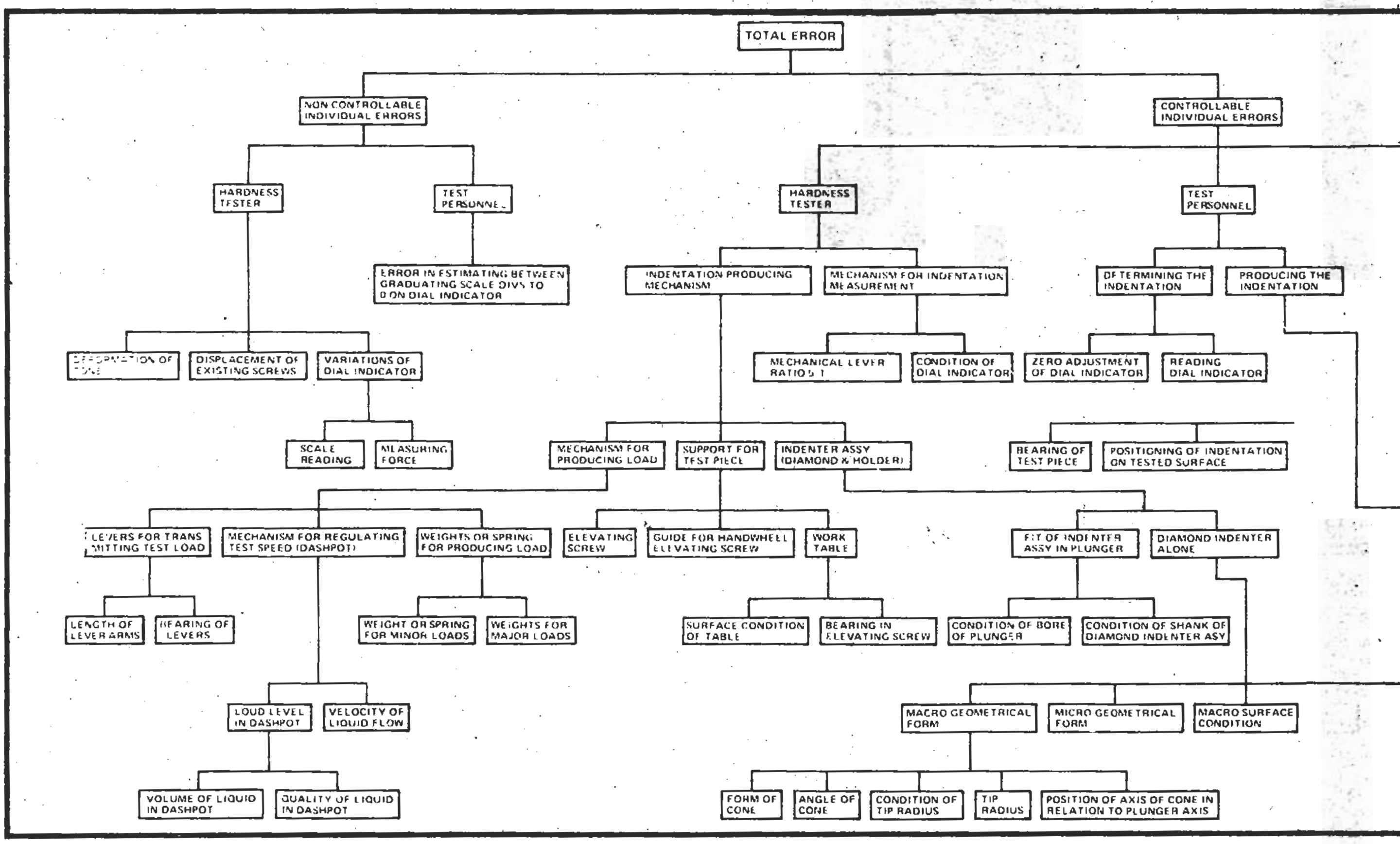
Resolution — at the finest scale available from a particular process of comparison — is the ability to resolve between successive increments in the chosen scale. A person using a mercury-thermometer might be able to resolve, say, 0.1°C intervals, subdividing these intervals into two divisions by eye gives a resolution of 0.05°C. By adding an optical magnifying system (or mechanical gain in, say, a micrometer — as depicted in Fig. 2) which has an inscribed scale at its focal plane, it is possible to raise the resolution 10 times or higher.

The unexperienced, unenlightened measurer will often quote this fact as a measure of how good a device is for measuring with but having resolution may not mean much in reality. It is very much the first basic requirement of comparison, but if it lacks accuracy the answers can be quite wrong. For example, if the optical magnifier on the thermometer has a badly ruled scale or optical distortion each increment will not be equal. Furthermore, the thermometer is supposed to measure temperature but, in fact, pressure of the air or liquid around it will also cause the mercury to rise or fall a little.

Precision — Two men argue as to which is the better rifle-shooter. It is to be settled by a contest on the range. The standard of excellence is to place the shots into the bulls-eye of the target but the game places the contestants sufficiently far back from the target that this is not easy to achieve. If it were, and each man placed all shots in the bulls-eye, the only assessment made would be that both were equally good. This situation lacks resolution to discriminate between them; the range is increased to increase the resolution. Thus emerges an important second rule of measurement — *there must be adequate resolution to a measuring process or little will be learned from the measurement.*

Each fires his group of shots and the two sets are intercompared. The first thing to be seen is that one group lies in a smaller total enclosed area on the target — as shown in Fig. 3a — but (in our chosen case) none is in the bulls-eye. The other shooter, on the other hand, has shots which are contained in a much larger circle, with one actually in the bulls-eye, as shown in Fig. 3b. The argument begins as to who is the better shot. The better shooter is probably the first, not the second, for his precision, that is, his ability to keep on the same spot, is much better than the other person. *Precision then, as the third rule, is the measure of scatter of values obtained in a test.*

In this case the first shooter would



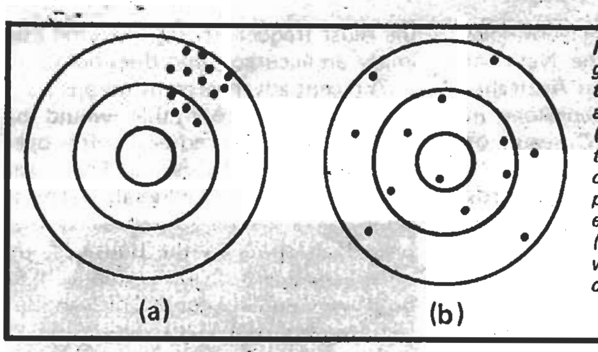


Fig. 3. Rifle-shooting is a good example illustrating the concepts of precision and accuracy.
(a) A precise group of shots that is not as accurate as could be — the mean position is not on the bulls-eye.
(b) An imprecise group which has an accurate mean centred on the bul'-eye.

be able to make a correction to his sight or allow for cross wind in order to move his group over to put more on the bulls-eye than the other person. The less precise shooter could not improve his accuracy. This leads us to what is accuracy in measurement.

Accuracy — In the shooting exercise the aim of each shooter was to reduce the distance between his individual shots and the bulls-eye — which is defined as the standard in this determination. The measurement that gives the closest value to the agreed standard is the most accurate. But in the example above the less precise shooter is, in fact, the most accurate if they decide that the averaged central position of the group is the criteria chosen.

In electrical terms a voltmeter may provide precise values but be very inaccurate due to a bent pointer or an altered value series ballast resistor.

Thus a fourth rule can be seen: *Precision and accuracy are quite different descriptive terms.* These are too often confused. Excellent precision does not imply equally fine accuracy and vice versa. It is always necessary to provide adequate resolution in order to determine the desired fineness of precision and to state the accuracy precisely enough.

TYPES OF ERRORS

The numerical value between the standard value and the measurand is termed the measurement error. Error magnitudes may affect the degree of precision and the accuracy obtained. They arise from many different sources, ranging from clearly identifiable processes, to never-identified mechanisms. Ideally, the measurer desires to eliminate all errors but the fact of life is that the closer we investigate a process in order to improve its resolution, precision and accuracy, the more errors loom up. Numerous sources of error can be identified with even the simplest of processes. Several years ago a study was made of the make-up of total error of a simple measurement (in principle at least) involving the pressing of a hardened point into a surface to measure its relative hardness by the degree of penetration — the Rockwell-C hardness test in this case. Something close to 40 sources of error were identified (as shown in Fig. 4).

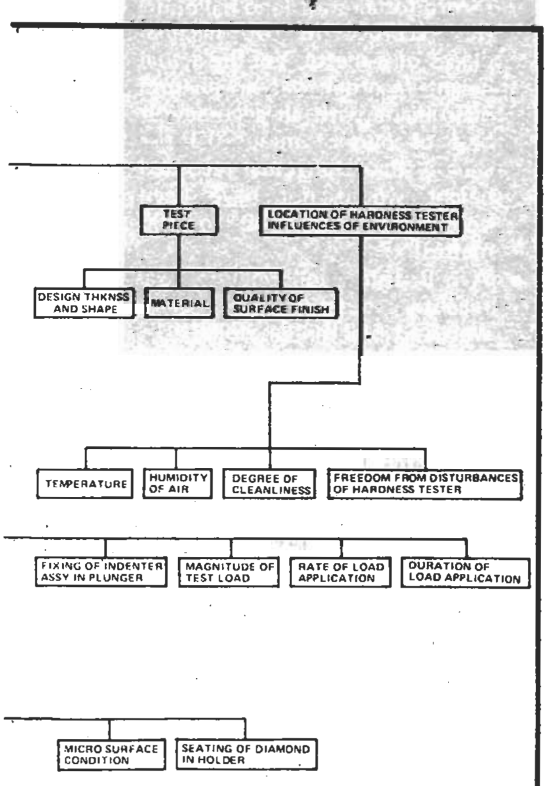


Fig. 4. Personal errors involved in using a simple technique of pressing a point into a surface to measure its hardness.

There are three main classes of error into which similar errors can be typified. Each has to be eliminated, reduced or lived with in different ways. A fifth rule of measurement is that *errors limit the usefulness of a measurement* and need to be reduced to tolerable levels.

Systematic errors — these are the derivations of values that always occur in the same way, and for which a corrective value can be applied to get the right value once the magnitude of the error is known. The rifle shooter resets his sights, the bent pointer is straightened. Or, the voltage reading can be corrected by adding the difference due to a bent pointer, or multiplied by a constant to make up for a wrong value series ballast resistor. It is, however, not necessary to know what *causes* the systematic error, only what its rules of occurrence are so that it can be allowed for.

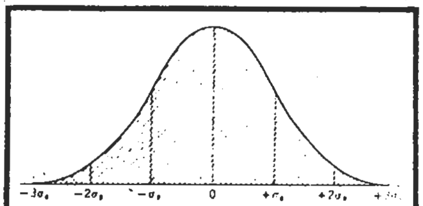


Fig. 5. Gaussian or top-hat distribution of random error.

Random errors — In strong contrast are errors that appear as the name implies, with random amplitude and sign. It is by definition impossible to predict what the random error will be on an individual value basis — the best that can be done is to place a level of probability of such and such a value arising at a certain time. In other words, seen as a group rather than a single occurrence of errors, it is possible to be reasonably certain about the value of such parameters as the *mean value* of the group and the *spread* of the group, but never the facts about the individual until it has occurred.

Random values follow statistical laws for collections of events. The most common occurrence of random error is with the so-called Gaussian distribution (also called top-hat or normal distribution). This form of error has a symmetrical profile for the plot of probability of occurrence of a value versus value changing as shown in Fig. 7. The peakiness of the curve is a measure of the spread of values and from the mathematical laws of this type or error it is possible to define a term that expresses the peakiness — the standard deviation (or s.d. or δ). In practical terms a s.d. of 1 means the limits $\pm 1\delta$ contain 68 per cent of values, $\pm 2\delta$ limits contain 95 per cent and $\pm 3\delta$ limits contain 99.7 per cent. If the chance of a value occurring is 50 per cent within a given limit and 50 per cent outside this limit then the probable error — has a value of 0.67δ.

It is conventional practice to quote the random error of a process in terms of the standard deviation, as this conveys the tightness of the random error in the measurement situation.

A trap that exists, however, is that not all random processes are Gaussian in distribution. Nuclear radiation particle occurrence, for instance, has a lop-sided distribution (Poisson) and another quite different set of mathematical formulae describes the chance of occurrence of values. In the majority of cases Gaussian statistics apply — white noise for example in electronic circuits.

Personal errors — To be correct these

Measurements

reduce to random or systematic causes of error, but by treating them as a specific group the deleterious effects of human observation are emphasised. There are numerous sources of personal error. The individual making the measurement may view a scale line at a different angle to another person thus introducing parallex error, the way of driving an adjustment screw dial may be different to another — moving from the opposite direction to the mark or at a different speed could introduce slightly differing values from one observer from another. In surveying practice it is quite normal for the theodolite or level operator to repeat the observation from the opposite direction. This reduces systematic errors of calibration by differential cancellation, thus reducing the personal error.

TRACEABILITY

We saw above how a standard must be created as the sole legitimate value of the unit, and how a measurement was made by comparing the unknown against this.

If the standard varies, then so does the measurement value. In the case of physical standards such as length, mass and time it is possible to provide some defined physical apparatus that acts as the standard. In some disciplines this is not so easy. Biological experiments use a control group — a group of test subjects that do not undergo the test given to the test group — as the temporary standard. In economic studies even the concept of a control group is hard to create, for we cannot ask half the country's population to stay the same economically and isolated at the same time as the other half have their financial situation altered. We would probably learn more about economic procedures if we could!

Whatever the standard it must be adequately constant for a long enough duration and be usable. For physical standards the demand for use is so great that it is necessary to have a sole fundamental standard controlling many working standards in each country, these controlling, in turn, the field standards used by individual laboratories. These control the value of the unit actually used in practice. Thus we can have as many as five or six steps between the fundamental standard and the working instrument. Each stage loses some accuracy so international standards must be maintained in the highest state possible by a national laboratory devoted to this task — the National Measurement Laboratory in Australia, the National Physical Laboratory in Britain, the National Bureau of Standards in the USA, etc.

Clearly if this tree of standards were not strictly controlled any individual unpoliced link could upset the sequence. The process of traceability is used to ensure that a measurement (at least with high-performance instrumentation) is traceable right through to the fundamental standard with the loss in accuracy being designed at each level. This concept is vital to the maintenance of standards in practical use.

MIRTH

Looking back in time, man's measurement endeavours include some highly amusing methods of producing standards. One early standard of length for the inch was 'three barleycorns, round and dry'. Another standard of length was prescribed by taking the first 16 men as they came out of church, making them stand with their left feet end to end — this gave a standard 'rod'. Not quite as bad as it may seem for at least a vaguely reproducible average was obtained. But the last man out defined the foot! In 1800, in Germany, there were 112 different size standards used to define just one common unit of length.

A peculiarity still with us today, concerns the gallon — the US gallon being smaller than the British Imperial gallon. In fact the US gallon is the earlier British gallon — the Pilgrim Fathers used the then-smaller Imperial gallon when they emigrated to the Americas in the 15th century. The Americans retained the original standard (more or less) but the British one was subsequently re-defined.

A 14th century treatise related an English penny — called Sterling — as the same weight as 32 grains of wheat. Thus in a very round-about way 20 pence make an ounce, 12 ounces a pound, eight pounds make a gallon of wine and eight gallons make a bushel of London.

Today some of our basic standards are still based on physical apparatus that is subject to damage or change. The standard (prototype as it is called) kilogram is still a piece of metal held in Paris. Most standards, however, can now be reproduced from a stated description of an apparatus which can be used to replicate the standard to within extraordinarily fine limits. Length for instance is defined as so many wavelengths of radiation from a Krypton discharge lamp.

Some extraordinary anomalies in measurement occur in every-day life — the most frequent being of a kind that imply an accuracy that does not exist.

A recent advertisement for a certain make of car — one that would be expected to be more careful over advertisements, says: "The car responds as quick as adrenalin". This is an entirely meaningless expression. It states nothing of substance. The same advertisement states that the car is "20 per cent safer than the safest car on the road". How do we measure safety in quantitative number terms? It is also said to have "precise rack and pinion steering" — let's hope so! And later: "Every one of its over 5000 parts is the result of adaption and re-adaption of . . . s" pioneering safety programme. It is very doubtful if every single one has been such — if so the designers need sacking for never getting there first-time in their design. Finally, "you need greater reserves of power to outdistance danger" — what a meaningless jumble of measurement statements! ●

- A recipe in a recent issue of an Australian magazine dutifully translated 'take 5 oz of flour' as 'take 141.75 grams' and half a pint of milk as '.354 litres'.
- Motoring magazines frequently quote standing quarter mile (or 400 metre) acceleration runs to three decimal places of one second. Yet one overseas magazine to my certain knowledge measures the required distance simply by a member of the staff pacing it out!
- A 'hundred thousand ton' ship was recently described in a daily paper as displacing '101,606.44 kilograms' — leaving aside that this contained an error of several orders of magnitude — the conversion implied that the original displacement was known almost exactly.
- After hearing that an aircraft was 'one minute late' I'm still trying to determine at precisely which point in its journey that an aircraft officially 'arrives'.
- Until very recently the altitude record for aircraft (and balloons) was recorded to two decimal places of a metre. Yet the actual height recorder was an aneroid instrument with an accuracy at best of plus or minus 0.1 per cent — thus the actual recorded height would not have been known to within 50 to 100 metres!
- A British millionaire was recently described as being worth $1.612 million dollars!