Measurement, Meaning and Morality

John Michael Linacre, Ph.D.
University of Sydney, Australia

Pacific Rim Objective Measurement Symposium (PROMS) &
International Symposium On Measurement & Evaluation (ISME), 2005

The modern perception is that science defines measures, which are then imposed upon the apparatus of society, in particular, commerce and politics. The metric system and the ratification in 1875 of the international "Treaty of the Meter," appear to bear this out. But history is actually far different. Measurement developed from the needs of practical persons and the impositions of their masters, who were the kings and rulers of ancient times. Modern educational and psychological measurement has the same motivations.

Measurement:
What is measurement? "Size or quantity obtained by measuring" (Pocket Oxford Dictionary). This is a definition by operation. It says there is an operation "measuring", which reports a "quantity". And what is a "quantity"? "A measurable or numerable amount" (Webster's New Collegiate Dictionary.) We can follow further along this chain of definitions, or consider the use of measurement in commerce, carpentry and cooking, and the conclusion is the same "measurement is the imposition of the rules of arithmetic on the world around us."

An essential precondition to the standard rules of arithmetic is that "one more unit means the same amount extra, no matter how much we already have." Of course, the arithmetic of real life only approximates this. "One more orange" is always "one more orange" when counting oranges, but is certainly not always "one more unit of orange juice". According to the Texas Department of Agriculture (2003), the commercially required range is 59cc. to 177 cc. of juice per orange fruit. Consequently, for commercial purposes, oranges are traded by weight or volume.

But even when we are apparently scientifically numerating exactly what we intend to measure, there can be confusion. The earthquake which shook Kamchatka in 1952 was of reported size 9.0 on the Richter scale. An earthquake of reported size 8.0 struck northern Japan on Sept. 26, 2003. The earthquake which leveled San Francisco in 1904 was 7.0 on the Richter scale. But what is the relationship between 7.0, 8.0 and 9.0? Does 9.0-8.0 = 8.0-7.0? No! The Richter scale is logarithmic: $10^{9.0-8.0} = 10^{8.0-7.0}$ so, when expressed in terms of simple arithmetic, the increase in earthquake "ground shake" power from 7.0 to 8.0 is the same as the increase from 8.0 to 8.28. An earthquake of twice the power of an 8.0 is of size 8.30. This is can be confusing.

So, whenever possible, commerce and science are conducted in additive units.

But what are additive units in psychology and education, in the unseen world of the mind? The "Final Report of the Committee appointed [by the British Association for the Advancement of Science] to consider and report upon the possibility of Quantitative Estimates of Sensory Events" concludes that measurement, in any true sense, is impossible in psychology, but "their opinion might change if new facts were established" (Ferguson et al., 1940). Happily, the new facts were established in 1953 by Danish mathematician Georg Rasch (Rasch, 1960).

There was an earlier parallel in astronomy. French philosopher Auguste Comte wrote: "*... the stars .... we would never by any means investigate their chemical composition....*" ("nous ne saurions jamais étudier par aucun moyen leur composition chimique", 1842), because he believed that such an investigation would require direct contact with a star. But direct contact is not required. In 1859, German physicist Gustav Kirchhoff used Joseph von Fraunhofer's newly devised technique of spectral analysis to determine the chemical composition of our nearest star, the Sun.

Establishing the new facts of measurement

Auguste Comte believed that, to obtain the chemical composition of a star, one must obtain a physical sample of it. Norman Campbell, the most influential co-author of the Final Report, believed that to measure, one must be able to perform a physical operation, a concatenation, such as placing rods end to end to measure length or piling bricks one on top of another to measure weight (Campbell, 1919).

But just as the composition of a star can be obtained indirectly by spectral analysis, social science measurement can be performed indirectly by probabilistic inference. Let us perform this apparent scientific impossibility ourselves.

Campbell's concatenation follows strict rules. When a rod of length X units is "added" (by performing a precisely defined operation) to a rod of length Y units, then their combined length satisfies the relationship:

$$\text{Rod}(X) + \text{Rod}(Y) = \text{Rod}(X+Y)$$

so that the arithmetical operation on the numbers X and Y concurs with a physical concatenation of rods of length X and Y.

In educational or psychological measurement, what happens when two persons, *m* and *n* encounter item *i*. Let them be of abilities B_{ni} and B_{mi} relative to item *i* expressed on an infinite linear latent variable. What happens when they respond to the item, individually and together? For Campbell concatenation, we require that:

$$\text{Outcome}(B_{ni}) + \text{Outcome}(B_{mi}) = \text{Outcome}(B_{ni} + B_{mi})$$

What is an outcome? Consider dichotomous items. In concrete terms, it is an observation of success or failure. But this is merely an instance - like shooting an arrow at a target. If the same process is repeated many times, or by many different persons of identical abilities on many different items of identical difficulties, then every observation is not expected to be the same. Rather there will be different responses with different frequencies. The frequencies represent past observations. They provide the basis for estimating the probabilities on which inferences beyond those observations are based. So the outcomes that are of interest to us are based on probabilities. But what, if any, function of probabilities supports the desired mental concatenation?

Let P_{ni} be the probability that person *n* succeeds on item *i*, then, by the definition of probability,

$$0 \le P_{ni} \le 1$$

This range does not correspond to the conceptually infinite latent variable so, transforming first with into an odds-ratio,

$$0 \le P_{ni} / (1-P_{ni}) \le \infty$$

and then into a log-odds-ratio,

$$-\infty \le \log (P_{ni} / (1-P_{ni})) \le \infty$$

Thus a log-odds transformation makes probability commensurate with an infinite latent variable. So, let us make the initial conjecture that the "Outcome" operation is arithmetically a log-odds transformation. Then

$$\text{Outcome}(B_{ni}) \equiv \log (P_{ni} / (1-P_{ni}))$$

and

$$\text{Outcome}(B_{mi}) \equiv \log (P_{mi} / (1-P_{mi}))$$

summing,

$$\log (P_{ni} / (1-P_{ni})) + \log (P_{mi} / (1-P_{mi})) = \log (P_{ni} * P_{mi} / ((1-P_{ni})*(1-P_{mi})))$$

[Here there are two people responding independently to the same dichotomous item at the same time. So their two responses can coincide (both right or both wrong) or not coincide (one right and one wrong). If we apply this logic to one person responding independently to the same dichotomous item at the same time. Then that person's one response always coincides with itself, i.e., self-coincides. The same situation arises with factorials: 3! = 6, 2! = 2, 1! = 1 so what does 0! equal? We find it convenient to conceptualize 0! = 1. Similarly, we can write the probability statement for when 3 people agree independently on a response ("coincide"), and 2 people agree independently ("coincide"), so what does it look like when one persons agrees independently on a response ("self-coincide")?]

This suggests a refinement to the "Outcome" conjecture. Let us now define "Outcome" to mean "log-odds of coincident scored observations". Then person *m* is always in self-coincidence so Outcome(B_{ni}) remains log (P_{ni} / (1-P_{ni})), and similarly for person *m* so that Outcome(B_{mi}) remains log (P_{mi} / (1-P_{mi})).

When person *m* and *n* work on the same item independently, the probability that they will both select a correct response category, and so succeed, is $P_{ni}*P_{mi}$. The probability that they will both select an incorrect response category, and so fail, is (1-P_{ni})*(1-P_{mi}). Non-coincidence is not allowed under the concatenation rule, so that segment of the sample space is ignored. Therefore

$$\text{Outcome}(B_{ni}+B_{mi}) = \log(\text{joint success / joint failure}) = \log(P_{ni}*P_{mi} / ((1-P_{ni})*(1-P_{mi})))$$

Thus, combining equations:

$$\text{Outcome}(B_{ni}) + \text{Outcome}(B_{mi}) = \log(P_{ni}/(1-P_{ni})) + \log(P_{mi}/(1-P_{mi}))$$
$$= \log(P_{ni}*P_{mi} / ((1-P_{ni})*(1-P_{mi}))) = \text{Outcome}(B_{ni}+B_{mi})$$

which is the relationship required for Campbell concatenation. So, through the device of probabilistic inference, we have concatenated the non-physical and discovered "new facts" which eluded the British Committee.

We can go further. B_{ni} is the ability of person *n* relative to the difficulty of item *i*, so let us define Outcome (B_{ni}) to be the difference between the ability B_n of person *n* and difficulty D_i of item *i.* Then

$$B_n - D_i = \text{Outcome}(B_{ni}) = \log(P_{ni} / (1-P_{ni}))$$

and

$$B_m - D_i = \text{Outcome}(B_{mi}) = \log(P_{mi} / (1-P_{mi}))$$

subtracting,

$$B_n - B_m = \log(P_{ni} / (1-P_{ni})) - \log(P_{mi} / (1-P_{mi}))$$

Examine this equation. The left-hand side is the difference between two abilities, independent of the item on which they are compared. The right-hand side is an expression of probabilities. These are inferred from observed frequencies in the data, and do not require knowledge of the specific difficulty of item *i.* In fact, item *i* can be any item. In other words, this relationship shows that, when concatenation holds, the difference between abilities is not only independent of the difficulty of the item used for the comparison, but the difficulty of that item need not be known. It was a similar comparison of item difficulties which caused Nobel-prize-winning economist Ragnar Frisch to exclaim "It [the person parameter] was eliminated, that is most interesting!" (Rasch, 1960, p. xviii).

The relationship equation

$$B_n - D_i = \log(P_{ni} / (1-P_{ni}))$$

is known as the *dichotomous Rasch model* and is conventionally written:

$$P_{ni} = \frac{e^{B_n - D_i}}{1 + e^{B_n - D_i}}$$

Thus not only the possibility, but also the specifics, have been obtained for constructing educational and psychological measures with the same measurement characteristics as physical measures. Any data with a probabilistic structure which accords with a Rasch model also accords with Campbell concatenation and so supports the estimation of measures which have the same arithmetical measurement properties as length and weight.

Measurement and Meaning

The numerical values of measures support commerce, politics and science. But the numbers themselves lack context. Consider a length of 100 units. Is it long or short? It depends on the context.

Let us consider the writings of Mohammed ibn-Musa al-Khowarizmi (830 CE), whose name originated "algorithm". He wrote Hisab al-jabr w'al muqabala (Calculating by restoring and comparing) known to most of us as "Algebra". He perceived that we want to find out about a "thing", which in Arabic is *shay'*, which to the Spanish sounded like *xay*, and which became known as *x*. So "x" is the unknown "thing". It is what we want to find out about. And we place it along the "x-axis". We call this the "latent variable". So, we want to place our measures meaningfully along the x-axis. We can then make a map, a picture of them. This gives them meaning. Here is such a picture:

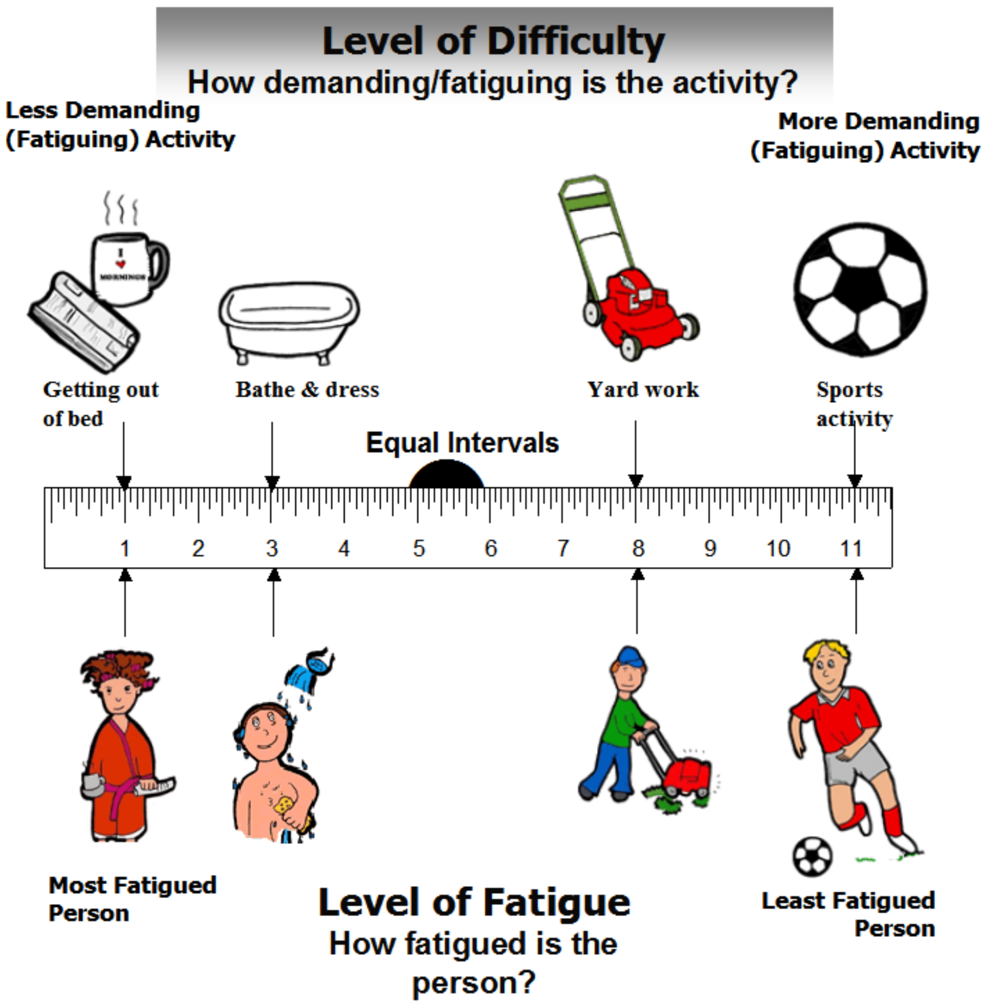

Figure: Measurement map of Fatigue. (Mallinson, 2001)

Now we know what a measure means both in terms of an item which exemplifies it (an agent) and a person who experiences it (an object). And here are some educational rulers.

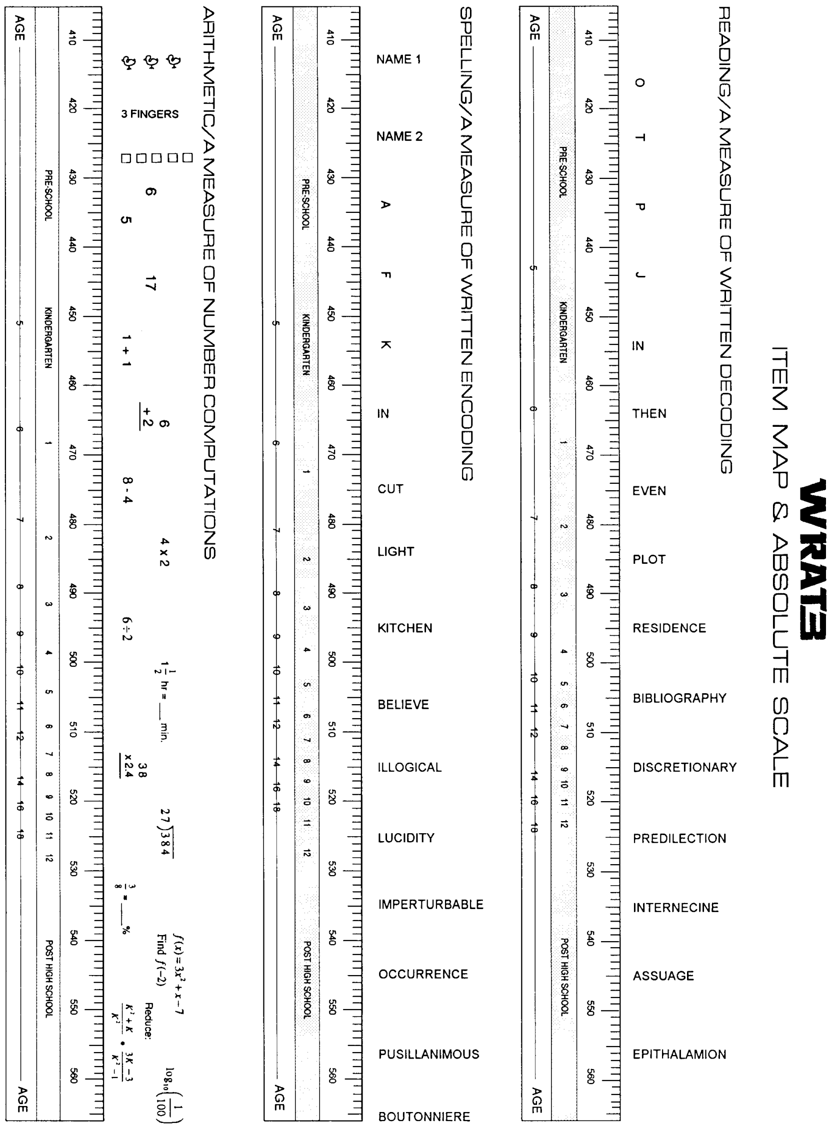

Figure. WRAT3 Item Map and Absolute Scale. (Jastak & Wilkinson, 1993)

When measures have context, they advance beyond merely being abstract numbers to having practical meaning.

Measurement, Meaning and Morality
The utility of additive measures for commerce, carpentry and cooking are familiar to us all, but the implications of honestly additive measures reach far beyond mere convenience.

Measurement must be ideal, but practical. Rigorous, but accommodating. Demanding, but forgiving. Quantitative, but qualitative. Forward-looking, but faithful to the past.

Science:
"The development of [physical] metrology from a great number arbitrary anthropomorphous measures to unified universal measures clearly shows that the same principle is being fulfilled in physics. The accord of ethical and physical principles was first noted by Sir Arthur Eddington, when in 1920 he chose the words from The Book of Deuteronomy* as an epigraph to the chapter on Weyl's unified theory in his "Space, Time and Gravitation" [Cambridge]. ... **We obviously live in**

the world where the fundamental principles of ethics and physics agree with each other." (Tomilin, 1999)

Politics:

謹權量 (jin quan liang)

"Chau conferred great gifts, and the good were enriched. ... **He carefully attended to the weights and measures,** examined the body of the laws, restored the discarded officers, and the good government of the kingdom took its course." (Confucius, The Analects, 20. ca. 500 BCE).

The political ramifications of not "attending to weights and measures" can be dire. Honesty is definitely the best policy. "One writer during the [French] Revolution felt it was no exaggeration to speak of sixty thousand measures of weight in France before 1789. ... it created a situation propitious for the falsification of standards by the *seigneurs* [feudal land owners] and for the distrust, justified or not, of the peasants. The old weights and measures upheld the old regime. A common demand of the *cahiers de doléances* [notebooks of grievances] of 1789 was thus to unify weights and measures - not to avoid paying feudal dues but to assure **an honest amount** payable. The rallying cry: *"un roi, une loi, un poids, et une mesure"* (one king, one law, one weight, and one measure) was a slogan of equality and centralization, the chief mark of modern French history, one that the monarchy commenced and the Revolution furthered." (Kennedy, 1989)

Religion:

وَيَا قَوْمِ أَوْفُوا الْمِكْيَالَ وَالْمِيزَانَ

"And O my people! give just measure and weight," (The Qur'an, The Prophet Hud, 11:85, ca. 600 CE).

Surely this instruction applies equally to educational and psychological measures as it does to physical measures. Measurement, meaning and morality - honestly, they work together.

References:

Campbell N.R. (1919) Physics: The Elements. Cambridge: Cambridge University Press.

Comte A. (1842) Cours de philosophie positive. (Course of Positive Philosophy). 6 vols. Paris.

Eddington A. (1920), Space, Time and Gravitation. Cambridge: Cambridge University Press.

Ferguson, A., Myers, C.S., Bartlett, R.J., Banister, H., Bartlett, F.C., Brown, W., Campbell, N.R., Craik, K.J.W., Drever, J., Guild, J., Houstoun, R.A., Irwin, J.O., Kaye, G.W.C., Philpott, S.J.F., Richardson, L.F., Shaxby, J.H., Smith, T., Thouless, R.H., & Tucker, W.S. (1940). Quantitative estimates of sensory events: Final report of the committee appointed to consider and report upon the possibility of quantitative estimates of sensory events. Advancement of Science, 1, 331–349.

Jastak S. & Wilkinson G. (1993) The Wide Range Achievement Test. WRAT3. Wilmington, Delaware: Jastak Assessment Systems

Kennedy E. (1989) A Cultural History of the French Revolution. New Haven: Yale Univ. Press. p. 77-8.

Mallinson T. (2001) Measuring the Impact of Fatigue on Everyday Activities during Chemotherapy. Paper presented at COMET, Chicago.

Rasch G. (1960, 1980) Probabilistic Models for Some Intelligence and Attainment Tests. Chicago: University of Chicago Press.

Texas Department of Agriculture (2003) Title 5. Chapter 94. Citrus fruit maturity standards.

Tomilin K. (1999) Natural Systems of Units. To the Centenary Anniversary of the Planck System. Proceedings Of The XXII Workshop On High Energy Physics And Field Theory. Protvino, Russia. http://dbserv.ihep.su/~pubs/tconf99/tomil.htm

* "You shall not have in your bag two kinds of weights, a large and a small. You shall not have in your house two kinds of measures, a large and a small. A full and just weight you shall have, a full and just measure you shall have" (Deuteronomy 25:13-15).