# Meaningful and Meaningless Statements Using Metrics for the Border Condition

Fred S. Roberts
*CCICADA Center*
*Rutgers University*
*Piscataway, New Jersey 08854*
*Email: froberts@dimacs.rutgers.edu*

***Abstract*—In the United States, there has been a search for metrics to measure the condition at the nation's border, with emphasis on border security. A goal is to produce a single metric or perhaps several metrics that could be used to assess whether the security at the border has improved or gotten worse. This paper considers metrics for the border condition and what claims using these metrics are meaningful, using a precise concept of meaningfulness from the theory of measurement.**

## 1. Measuring the Condition of the Border

In the United States, there has been considerable discussion since the early 2000s of the need for metrics to measure the condition at the nation's border, with emphasis on border security. A primary purpose of this discussion has been the goal of producing a metric or metrics that could be used to assess whether the security at the border has improved or gotten worse. A key aim is to be able to determine the effectiveness of prior investments in border security and to support decisions about future investments. There are some in the US Congress who have asked for a single metric that could be used to measure the border condition, and indeed there was a serious attempt by US Customs and Border Protection to develop a single metric called the Border Condition Index.

The problem of finding universally accepted metrics for the border condition is complicated for many reasons. These include the vastness of the US border and the numerous and different ports of entry for the legal movements of both goods and people, inbound and outbound. They also include the variety of transport modes that can be used to enter the country, e.g., on foot, by land vehicle, airplane or boat. One can argue that border security may depend upon activities far from the land border, which makes things even more complicated. Also, there are many agencies and other entities involved, each with different missions, objectives, and data collection processes. There are many components of border security, including the need to take into account the various "bad" things that need to be kept out of the country (narcotics, smuggled goods, illegal aliens, etc.), the need to develop border control policies that do not affect "good" commerce (by keeping inspection procedures from interfering, by cutting down on the length of time to obtain appropriate permissions and licenses, etc.), and the need to enhance or at least maintain the quality of life of those living at or near the border. For these various reasons, no universally accepted metrics have been achieved and indeed the idea that one can find one single metric (a border condition index) that captures all the components of the condition at the border has been challenged as impossible to achieve by many.

Ultimately, conveying border security is about decision making and the communication of information to policy makers and to the public. If developed properly, metrics can be used to convey information and to re-assure the public and policy makers. Metrics can be useful for this purpose, but they can also be misleading and statements using metrics can be downright meaningless in the precise sense of measurement theory.

It is critical to understand what conclusions might legitimately be drawn using any measures we develop. Do we want to say that the condition of the border has improved? We call this a *comparative statement*. Do we want to say that the improvement between years 2015 and 2016 was greater than it was between years 2014 and 2015? We call this a *comparison of differences statement*. Do we want to say that the condition of the border today is 10% better than the condition last year? We call this a *percentage change statement*.

The US Border Patrol breaks up its territory into sectors, which in turn are broken up into stations and in turn into zones. We might want to average border condition at different locations (different sectors, stations, zones) and make comparative statements, comparison of differences statements, and percentage change statements using these *average over sectors metrics*. What we want to say using metrics restricts the aggregation or averaging procedures we might be able to use to develop those metrics, as we will see.

Another approach we might consider is to develop metrics for different aspects of or criteria for border security, and then average these metrics in some way (e.g., by weighted average), to obtain one overall metric. Here we keep the sector fixed or report a value for the union of all sectors. One can ask the same questions as above for these kinds of averages, which we call *average over criteria metrics*.

We shall consider which kinds of claims about the border are meaningful under what conditions, using a precise concept of meaningfulness from the theory of measurement.

## 2. Meaningful and Meaningless Statements and the Border Condition

In this section, we present a brief introduction to the theory of meaningful statements and explore statements about border security in this context. Our approach is based on the ideas in [6]. In turn it is based on the notion, going back to the psychologist S.S. Stevens [9], [10], [11], that the properties of a scale are captured by studying *admissible transformations*, transformations that lead from one acceptable scale to another. For example, we transform temperature measurements from Centigrade into Fahrenheit by transforming $x$ into $\frac{9}{5}x$+32 and mass measurements from kilograms into pounds by transforming $x$ into $2.2x$. We will use the notation $f(a)$ for the scale value assigned by metric $f$ to $a$, the object or item or process or situation of interest. We assume that $f(a)$ is a real number. Then an admissible transformation of scale can be thought of as a function $\phi$ that takes $f(a)$ into $(\phi \circ f)(a)$.

Using the notion of admissible transformation, we can then characterize scales (metrics) by the class of admissible transformations. If the class consists of exactly those transformations of the form $\phi(x) = \alpha x$, $\alpha > 0$, we say we have a *ratio scale*; if the class consists of transformations of the form $\phi(x) = \alpha x + \beta$, $\alpha > 0$, we say we have an *interval scale*. Finally, if the class consists of all (strictly) monotone increasing transformations, we say we have an *ordinal scale*. Other scale types are defined in [6]. For ratio scales, the scale value is determined up to choice of a unit; in interval scales up to choice of unit and zero point; and in ordinal scales only up to order.

An idea going back to early work in the theory of measurement (see [12], [13]) is that a statement involving scales is *meaningful* if its truth or falsity is unchanged after any (or all) of the scales is transformed (independently?) by an admissible transformation. This reflects the desire that truth or falsity of a statement using scales not be an artifact of the particular scale values used.

In some practical examples in economics, psychophysics, and other fields (such as where the notion of semiorder applies), one has to use a somewhat different definition. See [6], [7] for a discussion.

Now, suppose we want to make a percentage change statement, e.g., to say that the condition at the border has improved by 10%. Then we are considering the statement $f(a) = 1.1f(b)$. It is easy to see that if $f$ is a ratio scale, then if this statement is true, it remains true if an admissible transformation (multiplication by a positive constant) is applied to $f$. However, it may not remain true if $f$ is an interval scale. Thus, for instance, it is meaningless to say that the temperature at $a$ is $10\%$ higher than the temperature at $b$. This could be an accident of the particular scale we are using (e.g., Centigrade or Fahrenheit). Thus, if our metric for the border condition only defines an interval scale, we will not be able to make claims about percentage of improvement.

Can we find a metric for the condition of the border that defines a ratio scale? We might for some components of border security. For example, under bad flows, we might measure number of kilos of cocaine interdicted. This defines a ratio scale. Similarly, one can count the number of illegal aliens captured. The numbers define what is called an *absolute scale*, a scale where the only admissible transformation is the identity. Of course percentage change statements are meaningful for absolute scales as they are for ratio scales.

However, if we want to find one metric that captures all the bad flows, we are faced with the dilemma of how to add kilos of cocaine to numbers of aliens, or how to translate each into commensurate values, let alone needing to add in other bad flows. Who knows if what you will end up with will define a ratio scale or even an interval scale?

To go even further, if we want to consider not interfering with good flows, we might calculate minutes of waiting time at the border, which defines a ratio scale. We might also want to calculate days of waiting time to get an import license renewed, which again defines a ratio scale. However, we have to combine both of these metrics and maybe others if we want a single metric for good flows. Measurement of quality of life at the border is more challenging. We might consider life expectancy (a ratio scale), number of years of education (a ratio scale), the length of working life (a ratio scale), severity of health disabilities, or the utility or value of life at the border. Severity of health disabilities is used by the United Nations [14] and it is not so obvious how to measure it, though most likely any useful metric will define an ordinal scale. Severity of cough is often used as a metric in medicine and frequently measured on a 5-point scale using the numbers 1 to 5 to indicate increasing cough severity. This defines an ordinal scale since only order matters. Another metric used in medicine is the Piper Fatigue Scale, which uses the numbers 1 to 10, again defining an ordinal scale. Utility functions are often thought to define interval scales (see for example [2], [3], [4].) It seems unlikely that we can combine all of these factors relevant to quality of life into one metric that is a ratio or interval scale, especially since at least one of the factors is only an ordinal scale to begin with.

Next, consider a comparison of differences statement, i.e., the statement that the difference between the scale values at $a$ and at $b$ is greater than the difference between those at $c$ and at $d$, i.e., $f(a) - f(b) > f(c) - f(d)$. This statement is invariant under admissible transformations of the form $\phi(x) = \alpha x + \beta$, $\alpha > 0$, so it is meaningful if $f$ is an interval scale or a ratio scale. It is not meaningful for ordinal scales. Thus, in terms of the border, we can meaningfully say that the improvement between years 2015 and 2016 was greater than it was between years 2014 and 2015 if our metric defines an interval scale, but not if it defines an ordinal scale. Thus, for example, we cannot meaningfully say that the difference in severity of health disabilities between 2015 and 2016 improved over the same difference between 2014 and 2015, but in many cases we can meaningfully say this for utility of life at the border.

Note that for interval scales, we can even meaningfully make percentage statements about differences, e.g., that the difference between 2015 and 2016 is 10% more than the difference between 2014 and 2015. This is because the statement $f(a) - f(b) = 1.1[f(c) - f(d)]$ is invariant under multiplication by positive $\alpha$ and addition of $\beta$.

While scales for individual factors of a component of border security like control of bad flows might define a ratio scale, it seems unlikely that we could combine them into a single metric that is even an interval scale.

For ordinal scales, we can meaningfully say that something improved, i.e., that $f(a) > f(b)$. That is, we can make comparative statements.

Here is a more subtle example. Suppose we report the date by which the year's target of captured cocaine is achieved. Suppose one year we reach the goal on July 19 and the next year we reach it on June 30. If the year starts January 1, then we had improved from 200 days to 180 days, or 10%. However, if the year starts October 1, then we had improved from 292 days to 272 days, an improvement of around 7%. Thus, it is meaningless to say that there was a 10% improvement from year to year in the date by which the goal had been achieved - unless the beginning date of the year is specified. The reason is that time defines an interval scale (allowing change of unit and of zero point) if we are talking date, but ratio scale (allowing only change of unit) if we are talking days, hours, minutes, etc. However, we can meaningfully say that the improvement in date we achieve a target between 2015 and 2016 is greater than it was between 2014 and 2015, since we are using (at least) an interval scale.

Suppose we want to average metrics, e.g., taken over different locations or over different criteria. For example, is it meaningful to say that the average scale value over sectors (or stations or zones) at time $t+1$ is greater than the average scale value over sectors at time $t$? Suppose that $a_i$ gives the scale value at sector $i$ at time $t+1$ and $b_i$ gives the value at sector $i$ at time $t$. Then we are looking at the statement

$$\frac{1}{n}\sum_{i=1}^{n} f(a_i) > \frac{1}{n}\sum_{i=1}^{n} f(b_i) \qquad (1)$$

This is a comparative statement using what we have called average over sectors metrics. Multiplication of all $f(a_i)$ and all $f(b_i)$ by a positive constant $\alpha$ does not change the truth or falsity of this statement. In other words, in case $f$ is a ratio scale, the statement that the average value at time $t+1$ is greater than the average value at time $t$ is meaningful. The statement is even meaningful if $f$ defines an interval scale. However, it is not meaningful if $f$ defines an ordinal scale. With ordinal scales, however, comparison of median values (as opposed to arithmetic mean averages) is meaningful. Thus, if we could develop an interval scale metric for the border condition, we could then make comparative statements with arithmetic mean averages over sectors. However, we have already expressed skepticism about the possibility of achieving an interval scale metric.

A similar analysis shows that comparison of differences using average over sectors is meaningful for ratio and interval scales, but not ordinal scales, and that percentage change statements using average over sectors is meaningful for ratio, but not interval or ordinal scales. For comparison of differences, we look at statements like

$$\frac{1}{n}\sum_{i=1}^{n} f(a_i) - \frac{1}{n}\sum_{i=1}^{n} f(b_i) > \frac{1}{n}\sum_{i=1}^{n} f(c_i) - \frac{1}{n}\sum_{i=1}^{n} f(d_i)$$

The situation is actually quite subtle. Let us consider a second way of averaging metrics of border security. Instead of averaging over sectors, let us fix one sector (or the union of all sectors), but consider different components or criteria for border security. We might for instance develop a metric $f_1$ for the ability to keep bad flows out, a metric $f_2$ for the ability to keep good flows moving, and a metric $f_3$ for the quality of life at the border. Then our overall metric could be a weighted average $M(a) = \frac{1}{n}\sum_{i=1}^{n} f_i(a)$ if we consider all of the criteria equally important, or

$M(a) = \sum_{i=1}^{n} \lambda_i f_i(a)$ if we weight the $i^{th}$ criterion by a weight $\lambda_i$. We have called this an average over criteria metric. Then if we want to say that the border index $M$ has improved, in the equal weight case we are considering the statement

$$\frac{1}{n}\sum_{i=1}^{n} f_i(a) > \frac{1}{n}\sum_{i=1}^{n} f_i(b), \tag{2}$$

where $a$ and $b$ are different times. If each $f_i$ is a ratio scale, we then ask whether or not (2) is equivalent to

$$\frac{1}{n}\sum_{i=1}^{n} \alpha f_i(a) > \frac{1}{n}\sum_{i=1}^{n} \alpha f_i(b),$$

$\alpha > 0$. This is clearly the case.

However, there is no reason to think that $f_1$, $f_2$, ..., $f_n$ have the same units. They might be entirely different scales with independent units. In this case, we want to allow independent admissible transformations of the $f_i$. Thus, we must consider

$$\frac{1}{n}\sum_{i=1}^{n} \alpha_i f_i(a) > \frac{1}{n}\sum_{i=1}^{n} \alpha_i f_i(b), \tag{3}$$

all $\alpha_i > 0$. It is easy to find $\alpha_i' s$ for which (2) holds but (3) fails. Thus, (2) is meaningless.

A similar analysis applies if we have a weighted average with unequal weights.

Does it make sense to consider different $\alpha_i$? It certainly does in some contexts. Consider the case where instead of sectors we evaluate animals and one criterion measures their weight gain while a second measures their height gain. In the case of the border, if one of the criteria is kilos of cocaine captured and another is minutes of wait time, we have a similar example.

So, even if average over sectors could lead to meaningful comparative statements, average over criteria might not. Even for ratio scales, we may not be able to combine them to obtain one single index by using averages of the scale values under different criteria if we want to be able to make even simple ordinal comparisons using the combined scales.

There is a large literature on various averaging procedures and the comparisons that may meaningfully be made using averages. See for example [1]. In general, you can always take weighted averages of metrics, since that is just an arithmetic calculation, but the question is what comparisons you can meaningfully make using such averages.

All of this is clearly relevant to the metrics used at the border, and suggests that great care must be taken if we are to take weighted averages of metrics developed to measure different components of the border security condition.

Even more subtle is the issue of what statistical tests one may make if we measure data on a ratio, interval, or ordinal scale. The application of ideas like those in this paper to descriptive statistics has been quite widely accepted; it was originally due to Stevens [9], [10], [11] and was popularized by Siegel in his well-known book [8]. We have principles like: Arithmetic means are ”appropriate” statistics for interval scales and medians for ordinal scales.

Over the years, Stevens’ ideas have also come to be applied to inferential statistics. They have led to such principles as the following: (1). Classical parametric tests (e.g., $t$-test, Pearson correlation, analysis of variance) are inappropriate for ordinal data.; they should be applied only to data which define an interval or ratio scale; (2). For ordinal scales, non-parametric tests (e.g., Mann-Whitney U, Kruskal-Wallis, Kendall’s tau) can be used. Not everyone agrees. This topic is beyond the scope of this paper. See [5] for a discussion of the issues.

**Acknowledgements.** The author thanks the US Department of Homeland Security for support under grant 2009-ST-061-CCI002-06. This paper has benefitted from discussions about metrics for border security with Ali Abbas, Dennis Egan, Nathaniel Heatwole, Paul Kantor, Isaac Maya, Curtis McGinity, Brian Roberts, and Henry Willis. In particular, parts of the introductory section borrow from a report written by Abbas, Kantor, Maya, and the author, with help of collaborators from the CCICADA and CREATE Centers.

## References


[1] Aczél, J., Roberts, F.S.: On the Possible Merging Functions. Mathematical Social Sciences 17, 205-243 (1989).

[2] Feldman, A.M., Serrano, R.: Welfare Economics and Social Choice Theory. Springer, New York (2006).

[3] Hunt, E.: The Mathematics of Behavior. Cambridge University Press, Cambridge, UK (2006).

[4] Pinto, C.A., Garvey, P.R.: Advanced Risk Analysis in Engineering Enterprise Systems. CRC Press, Boca Raton, FL (2012).

[5] Marcus-Roberts, H.M., Roberts, F.S.: Meaningless Statistics. J. of Educational Statistics 12 383-394 (1987).

[6] Roberts, F.S.: Measurement Theory, With Applications To Decisionmaking, Utility, And The Social Sciences. Reading, MA: Addison-Wesley (1979). Digital Reprinting Cambridge University Press, Cambridge, UK (2009).

[7] Roberts, F.S.: Limitations on Conclusions Using Scales of Measurement. In Pollock, S.M., Rothkopf, M.H., Barnett, A. (eds.) Operations Research and the Public Sector. Vol. 6 in Handbooks in Operations Research and Management Science pp. 621-671. North-Holland, Amsterdam (1994).

[8] Siegel, S.: Nonparametric Statistics, McGraw-Hill, New York (1956).

[9] Stevens, S.S.: On The Theory Of Scales Of Measurement. Science 103, 677-680 (1946).

[10] Stevens, S.S.: Mathematics, Measurement, And Psychophysics. In Stevens, S.S. (ed.) Handbook of Experimental Psychology, pp. 1-49. Wiley, New York (1951).

[11] Stevens, S.S.: Measurement, Psychophysics, And Utility. In Churchman, C.W., Ratoosh, P. (eds.) Measurement: Definitions and Theories, pp. 18-63. Wiley, New York (1959).

[12] Suppes, P.: Measurement, Empirical Meaningfulness And Three-Valued Logic. In Churchman, C.W., Ratoosh, P. (eds.) Measurement: Definitions and Theories, pp. 129-143. Wiley, New York (1959).

[13] Suppes, P., Zinnes, J.: Basic Measurement Theory. In Luce, R.D., Bush, R.R., Galanter, E. (eds.), Handbook of Mathematical Psychology. Vol. 1, pp. 1-76. Wiley, New York (1963).

[14] United Nations.: Handbook on Social Indicators, Series F, No. 9. New York (1989).