The status of the concept of reference object in measurement in the human sciences compared to the physical sciences

M Wilson¹, L Mari² and A Maul³

¹ Graduate School of Education, University of California, Berkeley, CA, USA
² School of Industrial Engineering, Università Cattaneo, Castellanza, VA, Italy
³ Gevirtz Graduate School of Education, Univ. of California, Santa Barbara, CA, USA

E-mail: MWilson@berkeley.edu

Abstract. For measurement in the human sciences, one barrier to being able to see the commonalities with measurement in the physical sciences has been the common prominence that is often given to the important place that “reference objects” are given in introductory accounts in metrology. The obvious difficulties in having such objects, say, “reference humans,” makes it difficult to build connections. In this paper, we outline a way to broaden the concept of a reference object in way that addresses this issue. In this light we then discuss the recent developments in the International System of Units (SI), and comment on certain similarities.

1. Introduction

Many basic introductory account of measurement feature centrally the idea of ‘reference object’ (also known as “measurement standards” in the tradition of physical metrology). For example, in [1], the initial step in his proposed formalization of measurement is to create a “standard reference set” (of objects) A, by associating them with a set of symbols Θ. The next step is then to carry out data acquisition: through empirical interaction, select an element of A that is indistinguishable from the object under measurement with respect to the considered property, and assign it the element of Θ that corresponds to the element of A.

When we look to the human sciences, this would seem to imply that we need to have accessible, stable and reproducible humans who we can (a) certify as reference objects, and therefore (b) use for comparisons in the practice of measurement. A technique for extending this logic to avoid the use of a set of human beings as reference objects has been discussed in [2]—the argument presented there does not attempt to detach from the primacy of the concept of reference objects, but instead builds on that concept to develop a different basis for measurement. Hence, this still assumes that at least in principle one could indeed find such “reference humans,” even though they are not required to be exploited in practice.

Given the common assumption that such reference objects are empirical entities such as rods, gauge blocks, standard cell electrochemical batteries, standard resistors, and so on, the typical pragmatic ploy in the human sciences is to replace the human reference object with a reference instrument. But, this ploy then leaves the reference object as missing in the context of measurement in the human sciences.
Thus, we address the question—what if there were a more direct relationship between the instrument and the human property that is under measurement? To see a different perspective than the typical one mentioned above, think about what we mean by “human being.” A human being is many-fold—incorporating a very large number of features, so many that they might (in practical terms) be thought of as infinite. This is one of the ways in which measurement in the human sciences is seen to differ from the stereotype of measurement in the physical sciences. In comparison, a rod used as a reference object for the property of length (say being a reference for a certain length—the meter etc.) would seem to have just one principal characteristic, the property of length. However, in fact, this is NOT true, for any such rod is actually also a cluster of a very large number of features. For example, each such rod is not mathematically straight, it has not the same width at every point, it does not have truly plane and parallel ends, it may not have the same electric charge, it does not weigh exactly the same as other “meter rods,” and it is not composed of the same atoms as other such rods, nor even the same mixture of types of atoms. Each meter rod used as a reference object is in fact uniquely itself, and different in many ways from every other reference object used to reference the meter. When this sort of example is being used in an introductory exegesis of measurement, these other properties are generally ignored, and hence a reader unfamiliar with physical measurement (such as even experts in measurement in the human sciences) can easily take away a stereotype that the reference objects are somehow “pure” examples of the property in question. Of course, it is an important point that, while physical reference objects can be designed to have one “principal” feature, that is not the case for human beings.

2. Our contentions
It is our first contention that we are held back from understanding the way in which instruments are indeed one aspect of reference objects by this underestimation of the individual diversity of the reference objects that are used in traditional measurement in the physical sciences. Or, put another way, we are over-appreciative of the individual diversity of human beings. This seems to us to be a forgivable weakness, as we are indeed all human beings, and we value our uniqueness both as individuals and as a species (at least until the artificial intelligences catch up with us!). Thus, a more humble perspective might well be that indeed a mere “meter rod” could be seen as having near-to the uniqueness of a human being. So, why do we not see a similar problem in creating reference objects for length as we do for human properties?

This introduces our second contention, that what we consider to be a “human property” is, in our thinking, overburdened with metaphysical complexities. For example, take reading comprehension ability as an example of a human property. When thinking about a reference object, we immediately imagine a person reading a text (an observable event), and, unobservably, forming a “comprehension” of the meaning of the text in their minds (this of course, is the erstwhile human property). This “comprehension” an observer might ascertain by asking certain questions of the reader, and interpreting/judging the responses. In fact, is it not a more straightforward stance to consider that the human property of reading comprehension is in fact inherent in the event of being asked questions (of a certain type, about the text and its content) and responding with (correct or incorrect) answers? From this standpoint, the property is seen as being an aspect of an event—the event of responding to the question(s).

3. Beyond reference objects
With this perspective in mind, one can then see that a reading comprehension test can indeed constitute a reference object (or, perhaps, more accurately, a system of reference objects). But the essence of such an insight lies in conceiving of the test as more than just the surface-level “object”—the test is not just the letters that form the words that make up the questions, or command, that constitute the items that make up the test. Each such item must be seen as the combination of (a) the prompt, (b) the response(s), and (c) the methods of interpreting those responses, in the sense of indicating the extent of the property revealed by
those responses. When an item is seen as being composed of these parts, it makes more sense to see it as (part of) a “reference object” for a human property that is wrought in a deeper way than would usually be conveyed by the term “object.” Hence, instead of relating measurement to “reference objects,” the contention here is that we can broaden the concept of an object, beyond being a “body” so that we could think of a “reference event.” In the context of reading comprehension, the event would be a reader reading a text passage, and responding to a series of questions posed about that reading passage. For each item, there would be a protocol for coding the responses to give a response category—the set of these for all the items would constitute the standard set of reference events, \( A \). And this would be used to generate a symbol from \( \Theta \) (as above).

Consider, for example a classic example of traditional measurement, the Kansas Silent Reading Test [2]. An example item from this test is:

\[
\text{I have red, green and yellow papers in my hand. If I place red and green papers on the chair, which color do I still have in my hand?} ______________
\]

This one might refer to as “nude” item—it stands alone, without reference to what it is purported to be measuring (apart from the title of the test, “Silent Reading”, it does not have specifications for scoring the item responses; it is assumed this is self-evident), and it does not come with a complicated procedure for accumulating the scores from the items to an overall test score (the scores are simply added up—as is common in education). This traditional style of item design encourages the conceptualization of each item as a stand-alone prompt, without the connections to a measurement system that would make clear the connection to the human event of engaging in certain type of situation designed to invoke a certain human characteristic, and responding in ways that can be then observed, coded and accumulated to make an inference about a person in terms of that property.

This can then be contrasted to an alternative example of an item that is indeed embedded in a context that scaffolds its status as a “reference event.” In this case, the theory about the person’s cognition takes a particularly simple form: the property is modeled as a “construct” consisting of a linear succession of discrete segments of a continuum, and when these are illustrated in a figure, it is termed a “construct map”. An example of this is shown in figure 1, which was designed as part of a test of middle school students’ practice of designing and critiquing displays of data.

![Figure 1. The levels of the Data Display construct map.](image-url)

Each item is developed to match to one or more levels of the construct map. An example item for the Data Display construct map is shown in figure 2. Figure 3 then shows sample student responses to this item that have been matched to certain levels of the DaD construct map. Gathering sample student responses can lead to a revision of the construct map and/or the hypothesized link, as well as modification/deletion of the item.

Once an item set is established, all items having the characteristics described in the previous paragraph, “banding” or “standard setting” takes place—this is the analogue of the “calibration” described by Mari in [1]: (a) locations for each item is estimated via a Rasch scaling, and then (b) the delineation of segments (“bands”) of the Rasch logit scale. The upper and lower boundaries of these bands are determined by the locations of the lowest and uppermost of the items linked to each level of the construct map. Quite often,
these item locations do not result immediately in a “clean” segmentation of the logit scale—hence an expert judgment process [3] is needed to gauge reasonable locations for the band boundaries of each band. This may require further decisions regarding the suitability of certain items, and may also involve demographic and other information about the persons, when that is available.

These successive bands can then be used as a basis for criterion-referenced interpretations of measurement results, enhancing and deepening the interpretations. At the same time, the existence of the underlying Rasch scale means that (a) technical aspects are available, such as estimates of instrumental uncertainty (standard errors etc.), and (b) technical advantages of item response scales are still available, such as flexibility in item choice, ability to link forms through items, and the possibility of computerized adaptive item administration.

Figure 2. An example item (“Shimmering Candle.”) related to the DaD construct.

Shimmering Candle Company claims that their candles burn longer, on average, than candles made by the Brilliant Candle Company. Testers for Consumer News burned 15 Shimmering candles and 15 Brilliant candles and recorded the number of minutes that each candle burned. The plot below shows the burning time and the middle-50 percent of the data around the median.

1. Is the claim made by the Shimmering Candle Company, that their candles burn longer, supported by the test results? Circle your answer.

   Yes    No

   Why do you think so?
Returning to the meter rod, we can see, from this broader perspective, that the rod itself, although it is indeed an object, is by and of itself insufficient for measurement. What is required is actually a reference event, which would be the comparison of that reference rod to the rod under measurement. And this must be carried out in certain ways (e.g., carefully aligning the “bottom” of the two rods, etc.), and under certain conditions (e.g., both being at the same temperature, etc.). And this would be coded by noting which rod was longer, or the same, and so on for the set of rods. And, as for reading comprehension, the assignment to $\Theta$ would be made on the basis of this set of codes.

![Figure 3](image-url)

**Figure 3.** Examples of responses to the DaD construct levels for the Shimmering Candles item.

Thus, this broadening of the specification of measurement basics not only makes the inclusion of typical human sciences practices more straightforward, but also makes the delineation of physical science measurement more comprehensive—making clear that it is not just the existence of a set of reference objects that matters, but also a set of informational specifications about how those objects must be used to produce the measurement result (i.e., part of the “measurement procedure”).

A person’s response to a set of items is certainly LESS than a human, but that can also be said of a rod—when we measure the rod by comparing it to a meter rod, we are ignoring all of the characteristics and features that make it an individual rod—we are only looking at one property—its length. And,
symmetrically, even though the meter rod is itself a rod, we are ignoring all of the characteristics and features that still pertain to even this “reference object.” It could as well be LESS than a rod, and it would still suffice for the purpose of being a reference object. In fact, one could say that this is a way to present measurement, and more generally science: the art of selecting something, and neglecting everything else in complex systems (i.e., “abstraction”).

4. Developments in physical sciences
In fact, this development is exactly what has happened, and continues to happen, in the advancement of measurement in the physical sciences—classical reference objects (even though they remain as a centerpiece of basic textbooks in measurement) have been superseded by events and even computations from theory—see for example [4]. Today the situation in physical sciences is radically different, even though we might still say that on the scene there are reference properties and reference objects. Now the SI units (a particular but important case of reference properties!) are defined as functions of properties of physical phenomena: the metre is the length of the path of light in vacuum in a given fraction of one second, and so on. By adopting this semiotic, encompassing concept of 'object', we may still consider units as properties of objects, but such objects are not physical bodies anymore: they are phenomena that are assumed to be invariant according to the best available theories (and the revised SI -- just approved by the CGPM -- makes this strategy of definition even more radical).

This situation in physical measurement is acknowledged in the VIM [5], where, as part of the definition of measurement standard—the "realization of the definition of a given quantity, with stated quantity value and associated measurement uncertainty, used as a reference"—the following is noted: "A “realization of the definition of a given quantity” can be provided by a measuring system, a material measure, or a reference material." Thus, the extension of the possible basis for measurement beyond the traditional measurement object (referred to as “measurement materials”) is encompassed within this broader definition. The case of a reference instrument, as discussed above would be included in the term “measurement system.”

5. Conclusion
The ultimate need remains to be able to calibrate measuring systems and therefore to guarantee the metrological traceability of measurement results, and since calibrating a measuring system against a phenomenon such as, say, light in vacuum, is not exactly a trivial thing to do, the definition of a unit is realized by producing (more traditional) objects against which (more easily) measuring systems can be calibrated.

Similarly, the extension of the conceptualization of measurement outlined here, replacing the basic concept of the reference object with a reference event, a practice common in the human sciences, can be seen as pointing the need for a broader conceptualization of measurement. One could speculate that what has happened is that the cementing of the stereotypical reference objects into the whole account of measurement has affected the way it is thought about and the way it is expressed. Then, when we come to a situation (such as measurement in human sciences) where the erstwhile reference object is less amenable to being used as such, the account has to be "turned around", and what is seen as being the first step may no longer be so--what is important is that we get to the same place in the end (after calibration).

From this perspective, the use of reference instruments in the human sciences should not seen as a sort of idiosyncratic step sideways due to the theoretical weakness of those sciences, but rather can be seen as a prescient advancement in the fundamentals of measurement that has been taken by certain of the human sciences to overcome the ontic complexities of human properties.
References

[1] Mari L 2000 Beyond the representational viewpoint: a new formalization of measurement *Measurement* 27 71-84

[2] Kelly E J 1916 The Kansas silent reading tests *J. of Educ. Psychol.* 7 63–80

[3] Wilson M and Draney K 2002 A technique for setting standards and maintaining them over time *Measurement and Multivariate Analysis* ed S Nishisato Y Baba H Bozdogan and K Kanefugi (Tokyo: Springer-Verlag) pp 325-332

[4] Zim X-Z 2018 The kilogram is dead. Long live the kilogram! *New York Times* November 16

[5] JCGM 2012 *JCGM 200:2012 International Vocabulary of Metrology—Basic and General Concepts and Associated Terms (VIM)* 3rd ed (2008 version with minor corrections) (Sèvres: Joint Committee for Guides in Metrology)