Measurement 51 (2014) 315–327

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

An introduction to the Rasch measurement approach for metrologists ☆

Luca Mari [a,*,1], Mark Wilson [b]

[a] School of Industrial Engineering, Università Cattaneo – LIUC, C.so Matteotti, 22, 21053 Castellanza (VA), Italy
[b] Graduate School of Education, University of California, Berkeley, 4415 Tolman Hall, Berkeley, CA 94720, USA

ARTICLE INFO

Article history:
Received 1 November 2013
Received in revised form 14 January 2014
Accepted 6 February 2014
Available online 19 February 2014

Keywords:
Metrology
Psychometrics
Rasch models
Measurement validity

ABSTRACT

In the interests of fostering an inter-disciplinary dialogue, increasing collaboration between "hard" and "soft" measurement scientists, and learning from one another, the paper develops an analytical discussion of common elements between metrology and psychometrics. A simple example of physical measurement is introduced according to the conceptualization and terminology of the *International Vocabulary of Metrology* (VIM), and then its structural analogy to a test using Guttman items is shown. On this ground the example is generalized so to include a probabilistic component, and this leads to the basic Rasch model. Some notes on the delicate issue of measurement validity conclude the paper, whose aim, in the long run, is a measurement-related shared concept system, and a terminology understandable in both physical and social sciences.

1. Introduction

Measurement science is a good context in which to consider, once again, the asymmetric relations between the natural and social sciences. The impressive effectiveness of their methods and instruments seems a sufficient reason for physicists, chemists, engineers, etc. to follow their path and be largely uninterested in developments in the measurement of non-physical properties. On the other hand, even though in many aspects emancipated from "physics envy", it is not unusual for social sciences to take physical measurement as a reference, and possibly a target point, given "their propensity to imitate as closely as possible the procedures of the brilliantly successful physical sciences" [1].

The fact that sensors implementing physical effects, the core components of physical measurement instrumentation, cannot be exploited for non-physical properties has discouraged passive imitation, and this has (at least in part) led to the development of different theories, methods, and instruments in the area of social measurement. The two disciplines grew up along parallel routes, sometimes approaching each other – a significant example is Finkelstein's endeavor to import representational theories in physical measurement – but also sometimes with clashes, as in the well known case of the committee activated by the British Association for the Advancement of Science in the 1930s (extensively discussed in [2]; a more concise analysis is in [3]), which produced, among other effects, the predominance of operationalism in psychological measurement for much of the 20th century, as well as Stevens' theory of scale types [4].

In the mentioned asymmetric situation, it may be interesting to continue exploring the contributions that social measurement has to offer to physical measurement. An

☆ A preliminary version of this paper was the basis for a keynote lecture jointly presented by the authors at the 2013 joint IMEKO TC1 – TC7 – TC13 symposium, 4–6 September 2013, Genova, Italy. Several enhancements resulted from the lively discussions at the symposium.

* Corresponding author. Tel.: +39 0331 5721.
E-mail addresses: lmari@liuc.it (L. Mari), MarkW@berkeley.edu (M. Wilson).

[1] One of the authors is a member of the Joint Committee on Guides in Metrology (JCGM) working group 2 (VIM). The opinion expressed in this paper does not necessarily represent the view of this working group.

http://dx.doi.org/10.1016/j.measurement.2014.02.014
0263-2241/

excellent context in which to pursue this goal is so-called *Rasch measurement*, an approach to measurement developed within the social sciences that posits that the mathematical model of measurement is such that:

1. The result of the experimental stage of measurement, i.e., the indication, is given in probabilistic, instead of deterministic, terms.
2. Measurands can be meaningfully compared by their ratio.

It is surely not a new subject and of which several introductory texts exist. However, these texts, and the scientific papers describing this approach to social measurement, are, in general, not always easily readable by physical measurement researchers and practitioners. There are at least two reasons of this difficulty:

- The emphasis in these texts on measurability specified in terms of algebraic conditions, whereas physical measurement is a moving target on this matter (for example ordinal measurement is routinely accepted nowadays).[2]
- Some critical differences in the presentation of basic concepts and the related terms, so that for example "latent trait" (or "latent variable") and "manifest observation" are sometimes used in social measurement for "measurand" and "indication" respectively.

Even the expression "Rasch measurement" sounds peculiar in metrology, where names are typically given to measurement principles (e.g., Peltier effect) and measuring instruments (e.g., Bourdon pressure gauge), and "x measurement" is reserved to x = given quantity, as in "force measurement". Apart from historical reasons, a possible justification of the expression "Rasch measurement" is that it can be thought of as referring to a combination of a *measurement method* and some assumptions on the underlying *measurement principle*. For this reason we will adopt here the more appropriate "Rasch measurement models", or the simpler "Rasch models". But are Rasch models actually *measurement* models? In the last part of the paper this delicate question is considered.

While we do not necessarily expect that Rasch measurement models would be immediately useful for physical measurement, a common, well founded understanding on them might foster more fruitful relationships between physical and social measurement, towards a desirable shared concept system and related terminology. This is the underlying purpose of the present paper, which introduces the basics of Rasch models by systematically interpreting them in the conceptual and lexical framework of the *International Vocabulary of Metrology*, third edition (VIM) [7], a freely accessible document that may be consulted in parallel to this paper (the first occurrence of terms taken from the VIM is in italics, so to ease the search of the corresponding definitions in the VIM).

The paper can be read as an interdisciplinary exploration of the concept of (mathematical) *measurement model* – "mathematical relation among all quantities known to be involved in a measurement" according to the VIM – particularly when specialized as a *measurement function*, i.e., the function that formalizes the (inverse) behavior of the *sensor* at the core of the *measuring instrument*, and that produces *measured quantity values* when applied to *indication values* and possibly values of other quantities such as *corrections* and *influence quantities*. The fact that this is a purely structural characterization makes it applicable in principle to both physical and social instruments: Rasch models are indeed measurement models in this sense, where typically *indications* are outcomes of tests (e.g., in the form of number of correct answers) and *measurands* are properties such as attitudes, abilities, … of individuals. A whole family of models is termed after Rasch, all sharing this basic structure. In Section 4 the simplest of them will be presented. A by-product of the paper is then to show that a significant case of measurement in social sciences can be effectively spelled out in metrological terms. An admittedly simple, and somehow artificial, example of physical measurement will guide us to recognize the analogies between physical transducers and tests, as they can be understood as measuring instruments of Rasch models and psychometrics in general (to emphasize such analogies the symbols will be maintained to be consistent in the two cases, thus departing from the accustomed symbols in Rasch models). The conclusions drawn from this comparison will be devoted to the validation of measurand definitions/models, an issue that physical and social measurement usually approach with different strategies.

Our hope is that from what follows natural scientists and engineers may learn something of Rasch models, as a specifically relevant case of social measurement, and social scientists may re-interpret something of their knowledge of measurement in the light of the current physical measurement models.

2. Example 1: Hookean springs and Boolean springs

With the aim of measuring a given force f, a spring can be exploited as an *indicating measuring instrument*, and specifically as a sensor, which is supposed to behave according to a transduction function (sometimes also called "observation function") specified by Hooke's law:

$$x = \frac{f}{k} \tag{1}$$

i.e., the measurement principle is that a force f applied to a spring of elastic constant k generates an elongation x in the

[2] An interesting example concerns the possible requirement that the scale is continuous, or at least its elements are dense (i.e., isomorphic to rational numbers), that in Holder's axioms is expressed as "For every magnitude there exists one that is less" (and note that Holder himself presents his aim in this way: "I intend only to propose a simple system of axioms from which the properties of the *ordinary continuum of magnitudes* can be derived." – emphasis added) [5]. This implies that counting cannot be a type of measuring and that discrete properties are not measurable. According to Michell [2], "we have no good reason to suppose that measurable quantities are not continuous." This is in stark contrast to the views of many scientists. Consider, for example, the following quotation from Richard Cox, working from a different tradition: "reflection suggests, indeed, that the only perfectly precise measurement is counting and that the only quantities defined perfectly are those defined in terms of whole numbers" [6].