Measurement 92 (2016) 489–496

Contents lists available at ScienceDirect

Measurement

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

Theory-based metrological traceability in education: A reading measurement network

William P. Fisher Jr. [a,b,*], A. Jackson Stenner [c,d]

[a] LivingCapitalMetrics.com, Sausalito, CA, USA
[b] BEAR Center, Graduate School of Education, University of California, Berkeley, CA, USA
[c] MetaMetrics, Inc., Durham, NC, USA
[d] School of Education, University of North Carolina, Chapel Hill, NC, USA

ARTICLE INFO

Article history:
Received 14 October 2014
Received in revised form 11 April 2016
Accepted 20 June 2016
Available online 21 June 2016

Keywords:
Psychometrics
Reading ability
Education
Predictive theory
Unit standards
Test equating
Social studies of science
Commercial applications
Traceability

ABSTRACT

Huge resources are invested in metrology and standards in the natural sciences, engineering, and across a wide range of commercial technologies. Significant positive returns of human, social, environmental, and economic value on these investments have been sustained for decades. Proven methods for calibrating test and survey instruments in linear units are readily available, as are data- and theory-based methods for equating those instruments to a shared unit. Using these methods, metrological traceability is obtained in a variety of commercially available elementary and secondary English and Spanish language reading education programs in the U.S., Canada, Mexico, and Australia. Given established historical patterns, widespread routine reproduction of predicted text-based and instructional effects expressed in a common language and shared frame of reference may lead to significant developments in theory and practice. Opportunities for systematic implementations of teacher-driven lean thinking and continuous quality improvement methods may be of particular interest and value.

1. Introduction

Metrology connects measurement applications across industrial, scientific, and practical tasks separated by space and time. Significant fractions of many nations' economic productivity are invested in ensuring traceability to standards for various units of measurement. The human, social, environmental, and economic value of the returns on these investments depends on the transparency of the measures and their integration into a wide range of decision processes at multiple organizational levels. Huge resources are required to create and maintain technologically produced effects, such as volts, seconds, or meters, with the primary return on those resources being the illusion that the effects seem to be products of nothing but completely natural processes occurring with no human intervention.

New insights into how cognitive, social and technological resources aid in creating shared cultural frames of reference have emerged from close critical study of historical and contemporary scientific modelling and metrological practices. From this perspective, science is not qualitatively different from everyday ways of thinking and relating, except in more deliberately extending laboratory processes into the world as distributed cognitive systems supporting a range of associated problem-solving methods [1–5]. Of particular interest here is the linking of specific ways in which organizations align and coordinate their processes and relationships relative to technical developments and expectations. A positive result of adopting this point of view is recognition of the value of previously obscured accomplishments in, and opportunities for, advancing the quality of research and practice in psychology and the social sciences. An illustrative example is found in the scientific modelling and metrological practices informing integrated reading assessment and instruction in education.

1.1. Transparent instruments, invisible production

By definition, metrologists are doing their jobs best when no one knows they are there. Experimental scientists, for instance, may take little notice of their instrumentation until it breaks down or does not conform to expected standards. The general public and

* Corresponding author at: BEAR Center, Graduate School of Education, University of California, Berkeley, 2000 Center St, Suite 301, Berkeley, CA 94704, USA.
E-mail addresses: william@livingcapitalmetrics.com (W.P. Fisher Jr.), jstenner@lexile.com (A.J. Stenner).

http://dx.doi.org/10.1016/j.measurement.2016.06.036
0263-2241/

researchers in psychology and the social sciences are, then, also largely unaware of the resource-intensive work involved in establishing uniform unit standards and traceability to them [3,4,6,7].

The uniformity of the various phenomena described by natural laws allows scientists the convenient efficiency of not needing to specify scale units in statements of laws. Force equals mass times acceleration in kilograms, Newtons, and meters just as well as in pounds, poundals, and feet. The ability to skip over uniform details supports a division of labour in science that separates theoretical work from the calibration of instruments and both of these from the use of theory and instruments in experiments [8].

The convenience of separating theoretical, experimental and instrumental concerns has its drawbacks, too. Not knowing when or how reference standard units are established reinforces unexamined metaphysical assumptions—such as the idea that the universe or nature is inherently and innately numerical, quantitative, or mathematical—that rarely become explicit objects of attention.

The effect of these presuppositions is significant. Huge social, industrial, and economic efficiencies are gained by universal consensus on the facts of complex phenomena like electricity, temperature, distance, mass, and time. Though the dynamics of that consensus are complex and sometimes counterintuitive [8], making quantities seem natural is a cultural achievement of the highest order.

The advancement of science is put at risk when the historic and historical mathematical understanding of scientific objects is reified as unquestioned and unquestionable. Two questions emerge here: (1) how did the natural sciences succeed in making quantities seem so thoroughly natural [3,4,9–12], and (2) how might the social sciences learn from those successes? Recent advances in reading measurement embody important lessons in this regard for the social sciences.

1.2. Shortsightedly focusing attention on the local measurement outcome

The technical processes of measurement were historically cut out of the picture of science by the positivist focus on empirical observation, as well as by the later anti-positivist focus on theoretical constraints on observation [8]. Sometimes this omission was literal and deliberate, as when a woodcut of a laboratory scene printed in its entirety in one place is trimmed in a later publication to exclude the means by which a technical effect was produced [12]. Other times the omission was metaphorical, as when technical processes were illustrated in summary form by angelic cherubs producing effects by means of divine intervention [12].

Transparency in measurement is a two-edged sword. Wide access to comparable measures is achieved only to the extent that technical complexities can be ignored. This point was emphasized by Whitehead [13], who observed that "Civilization advances by extending the number of important operations which we can perform without thinking about them" (p. 61). But what happens when those making these advances do not record—or do not themselves fully understand—*how* they extended the number of important operations that can be performed by persons unversed in their technicalities?

In his study of the geometric assumptions Galileo employed in his physics, Husserl [14] was sensitive to the ways in which a hidden agenda set priorities. Like Galileo, we find ourselves in a situation, in accord with the philosophical problems attending measurement, in general, where

> Metrology has not often been granted much historical significance. ... Intellectualist condescension distracts our attention from these everyday practices, from their technical staff, and from the work which makes results count outside laboratory walls [6].

Researchers in the natural sciences make use of commercially available precision tools calibrated to universally uniform reference standards, standards capitalizing on the value of invariant laws. Transparent measures communicated in a network sharing common values situates metrology's often unrecognized historical significance in a complex overall context offering important lessons for psychology and the social sciences [1–12]. The culture of science rewards a mix of convergent, divergent, and reflective thinking in ways that have proven their productivity and inform a vital culture of ongoing innovation [8,9,11,12,15].

1.3. Consequences for psychology and the social sciences

But in the social sciences, the lack of metrological institutions, methods, and traditions, and the associated absence of the intercalated disunity of distinct theoretical, experimental, and instrumental communities observed by Galison in the natural sciences [8], has been catastrophic. As social scientists have long recognized for themselves [16–18], mainstream research methods and statistical models employ scale-dependent ordinal data in a search for a kind of significance that is often irrelevant to and even antithetical to the production of new knowledge. Even when regularities akin to natural laws are sought and found in psychological and social phenomena [19–23], results are typically assessed in the language and methods of statistics rather than of measurement and metrology, meaning the focus is on data analysis and not on theory development or the calibration of instruments traceable to a standard unit. The human, social, economic, and scientific consequences of this failure to coordinate and balance convergent, divergent, and reflective field-defining activities are profound. Ideas on how such activities might be organized in education have recently been proposed [24].

The lack of institutions and traditions concerning metrological traceability and standards in psychology and the social sciences may have more to do with broad and deep cultural presuppositions than with an actual lack of a basis for them in evidence. After all, what systematic program of experimental evaluation has ever irrefutably established that uniform metrics based in lawful regularities are impossible in psychology and the social sciences? Evidence indicates that provisional possibilities exist in some circumstances [19–24].

2. Metrological traceability for reading measurement

The longstanding need to provide students with reading challenges appropriate to their reading abilities is usually approached in terms of general curricular structures, and teacher training and experience. Theory has not been of significant interest [25,26]. Rasch's development of a new class of measurement models in the 1950s was an important step forward in improving the quantification of reading ability [25]. This research led to improvements in the matching of readers to text.

When Rasch's concept of specific objectivity (the modelled independence of the ability and difficulty parameters, as shown in Eq. (1)) as it was obtained in local measures was combined with a general predictive theory of English text complexity in the 1980s, following the work of Stenner and colleagues [27–29], the stage was set for the efficient creation of a network of reading measurement instruments calibrated in a common unit. By the late 1990s, all of the major high stakes English reading tests in the U.S. had been brought into the system. These are today complemented by the hundreds of thousands of books, tens of millions of short