Metrology, psychometrics, and new horizons for innovation

William P. Fisher, Jr.1,*
1LivingCapitalMetrics.com, Sausalito, California 94965 USA, and BEAR Center, University of California, Berkeley, California, 94720 USA

Abstract. The vital roles of common languages, metrics, and standards in team-based innovation are well established. The historical success of science has depended extensively on mathematically rigorous, theoretically explained, experimentally evaluated, and socially distributed communications networks comprised of metrologically traceable measuring instruments. Recent descriptions of practical and theoretical correspondences between metrology and psychometrics extend scientific model-based reasoning into education and other fields. Realizing psychometrics’ metrological potentials requires attending to interrelated ethical, technical, and social issues. First, measurement must embody the ethics of the Golden Rule in support of compassionate and tolerant comparisons. Technically, effective measurement systems involve (a) invariant units, with known uncertainties, (b) scientific laws modelling substantive relationships in the world, (c) hypotheses experimentally tested against data, (d) explanatory theories providing predictive control over the construct, and (e) methods for ensuring traceability to the standard unit. But to be effective, measurement systems incorporating these features will need to be situated in a larger social context cognizant of needs for the relevant standards bodies, educational programs and textbooks, professional societies, legal and regulatory oversight, accounting and economic applications, etc. Paraphrasing Rasch, the challenges are huge, but having formulated the problem, human ingenuity will likely find a way to solve it.

1 Introduction

Over the last several decades, two (TC-1 on Metrology Education and TC-7 on Measurement Science), and then three (adding in TC-13 on Measures in Medicine and Biology), of the International Measurement Confederation (IMEKO) technical committees have held Joint Symposia on the human and social sides of measurement. These meetings take place in the two years between the tri-annual IMEKO World Congresses. The participants have largely been metrology engineers and physicists interested in the larger philosophical and theoretical issues of measurement and quantification, especially as concerns the possibility of establishing a common conceptual frame of reference across the

* Corresponding author: william@livingcapitalmetrics.com or wfisher@berkeley.edu
natural and social sciences [1-12]. There have also been significant contributions from medical laboratorians, educators, psychologists, economists, and others [13-15].

Since the 2008 IMEKO Joint Symposium in Annecy, France, participation of researchers in education and health care making use of Rasch’s probabilistic models of measurement has grown markedly [16-32]. A number of engineer-psychometrician collaborations have emerged from this work [33-39]. The role of psychometricians in the Joint Symposia has expanded, with the 2016 Joint Symposium hosted by the Berkeley Evaluation and Assessment Research (BEAR) Center in the Graduate School of Education at the University of California, Berkeley [40]. On this occasion, there were about equal numbers of engineers and psychometricians presenting and in attendance, with over 70 registrations. The 2017 Joint Symposium in Rio de Janeiro, Brazil, again has a number of psychometric presentations on the program, including a keynote from David Andrich [41], an early leader in theoretical and practical developments in Rasch measurement [42-45].

2 Psychometric metrology: Challenges and opportunities

Realization of psychometrics’ metrological potentials will require close attention to interrelated ethical, technical, and social issues. First and foremost, measurement must embody the ethics of the Golden Rule, treating others in a way that is sensitive not just to how I would want to be treated but to how I would want to be treated if I were them. The technical challenges associated with realizing a social ethic based on the Golden Rule must resolve the paradox of aiming “at two opposed things: human totality and human singularity” [46]. In accord with Ricoeur, it is essential to resolve the paradox by integrating the apparently opposed things at a higher level of metasystematic complexity [47]. Systems are typically conceived mechanically, such that aggregations are sums of parts, when what are needed are organic wholes greater than the sums of their parts. Technical and social solutions to this kind of multilevel problem take living language as their model [48], and are perhaps more accessible and feasible than is generally realized [24,36,38,49-51].

2.1 The Golden Rule

Compassionate and tolerant comparisons avoid committing the violence of the premature conclusion, as the philosopher Paul Ricoeur called it [52]. The Golden Rule [53] provides the model for proceeding. It requires some way of knowing that the way I treat others is equivalent to the way I would want to be treated. To know this, I need to be able to project a basis for comparison, a kind of virtual, qualitative measuring system that keeps things in proportion by showing what counts as the same thing for different people.

Just as the whole of many religious and philosophical approaches to life is captured in admonitions to not do to others what you would not want them to do to you [53], at root, all violence is the violence of the premature conclusion [52]. When prejudices, hegemonic agendas, fear, impatience, greed, etc. dominate relationships, we leap to unjustified conclusions and unfairly reduce others to caricatures of what they really are. This kind of reductionism and imposition of power is a choice in favor of violence over discourse. But if we choose discourse and the possibility of keeping the conversation going, then we have to stay open to seeking better representations of others and their positions. Of course, though we can remain open in principle to new information, it will be necessary at some point to make a decision. And so things are complicated when we stop short of physical violence, but then shut down the conversation in various ways, preventing others from representing their social, economic, and political interests under the uniform rule of law.
So, recognizing that our judgments are always provisional and that no sample of evidence is ever absolutely conclusive, we have to be able to tell when our evidence is sufficient to the task of representing where someone stands. The technical challenges may be intimidating, but the sufficiency and necessity of Rasch’s models [45,54] for integrating information across levels of complexity [50,55] is well established. The social challenges will likely be more difficult to overcome than the technical ones.

2.2 Technical considerations

To apply the Golden Rule there has to be some way of knowing when a service or action provided to others is the same as the one provided to me. The Golden Rule could then be enacted via comparisons based in analogies between people's values and what is valued. The usual way in which measurement information systems are conceived involves imposing rigid criteria onto situations to make them adhere to some kind of a law [48].

That is, systems for distributing and supporting the communication of measurement information have not typically taken into account all-pervasive discontinuities between denotative, metalinguistic, and metacommunicative levels of discourse [51,56]. To be able to say that my values are to one aspect of a situation as yours are to that or another aspect requires proportions of this kind to hold constant, within a known range of uncertainty, no matter which particular persons are addressed and no matter which aspects of the situation are involved. Furthermore, to be able to maintain the meaningfulness of these kinds of comparisons across levels of complexity, individual-level responses must organize themselves into repeatable invariant patterns at a group level, patterns that cannot be imposed from outside of the self-organizing process and that are not intended by any individual. The constructs measured and managed via Rasch models are these kinds of self-organized group-level phenomena, opening the door to potential resolutions of Ricoeur’s paradox of human totality and human singularity.

These kinds of repeatable, invariant, and proportionate relationships are built into everyday model-based reasoning and have been extended into science via processes of analogy and metaphor [57-63]. The scientific status of psychology and the social sciences has long been conceived as contingent on shallow imitations of the methods and mathematics used in the natural sciences. Beginning from the routine cognitive processes used by humanity in general offers a far more promising path for enhanced creativity and innovation in these fields. And so it happens that Rasch [64] employed, in a limited and incomplete way [25,65], the method of analogy devised by Maxwell that was crucial to his successes in thermodynamics and electric field theory [57].

Statistically sufficient, invariant, and separable model parameters are the fundamental requirement of Rasch’s probabilistic models for unidimensional measurement [43,44,64,66-71]. The probabilistic form of psychometric models of measurement facilitates development of new ways of keeping the question open while still supporting decision processes and comparability. Provisional ways of not leaping to premature conclusions are built into probabilistic approaches to measurement in that survey, assessment, or test questions can be added or dropped, or adaptively administered, without compromising the comparability of the unit of measurement [72].

Technically, effective measurement systems involve (a) invariant units, with known uncertainties, (b) scientific laws modelling substantive relationships in the world, (c) hypotheses experimentally tested against data, (d) explanatory theories providing predictive control over the construct, and (e) methods for ensuring traceability to the standard unit. Information systems in education, health care, and many other fields incorporating measures have been available and in use for decades. Incomplete conceptualization of the social processes by which measures become integrated into everyday routines [73] have
hampered the fulfilment of the potential these systems have for resolving many outstanding issues of our time.

2.3 Social issues

Alder [74] points out that
To do their job, standards must operate as a set of shared assumptions, the unexamined background against which we strike agreements and make distinctions. So it is not surprising that we take measurement for granted and consider it banal. Yet the use a society makes of its measures expresses its sense of fair dealing. That is why the balance scale is a widespread symbol of justice. ... Our methods of measurement define who we are and what we value.

He continues, saying that
Just as the French Revolution had proclaimed universal rights for all people, the savants argued, so too should it proclaim universal measures.

Ricoeur [75] explains how it happens that standardized symbol systems inform the background assumptions that make social life possible:
...the function of substituting signs for things and of representing things by the means of signs, appears to be more than a mere effect in social life. It is its very foundation.
We should have to say...not only that the symbolic function is social, but that social reality is fundamentally symbolic.

Science extends everyday language and model-based reasoning by systematically using them to format processes and outcomes in ways that make them manageable and communicable [57,58]. This is not a new idea. Huxley wrote in 1862 that “The method of scientific investigation is nothing but the expression of the necessary mode of working of the human mind” [76]. Einstein similarly held that "The whole of science is nothing more than a refinement of everyday thinking" [77]. Natural language’s cognitive ecosystems provide the model for scientific language’s metrology networks [80-82]. Duncan accordingly [83] comments that “All measurement is...social measurement. Physical measures are made for social purposes” and that “The social roots of social measurement are in the social process itself.”

The questions that arise in this context concern how the social processes giving rise to metrological traceability and unit standards can be extended into psychology and the social sciences. We have had in hand in psychology for decades mathematical proofs of sufficiency [45,54], experimental evidence of invariance [84], explanatory causal theories [85], and networks of instruments measuring in common units [26]. So why have economists and accountants been proposing models of human, social, natural, and manufactured capital for years—as alternatives to the traditional models of land, labour, and manufactured capital for years—as alternatives to the traditional models of land, labour, and manufactured capital [86-88]—with no concomitant interest in the scientific measures needed to manage those stocks and flows? Why are the concepts of private property—and the capacity to profit from investments in it—not yet extended to the various forms of human, social, and natural capital? With over 50 years of successful high quality measurement of literacy, health, social, community, and natural capital, why does research into these forms of capital [89-93] not even raise the question as to if or how common currencies for the exchange of value might be devised?

How can we have universal human rights without universal measures of most forms of the capital needed for a functioning economy? Why is it impossible to know how many shares of literacy or health capital stocks we own, or what they are worth? Why are there no legally defensible and financially fungible forms of legal title to those shares? Why are investments in social capital not accountable to those providing the funds? Why must philanthropic capital markets remain so dysfunctional [94]?
3 Conclusion

Answers to these questions and widely use of alternatives to existing mainstream psychometric methods and systems will emerge only when a science of precision measures connected via metrological traceability is complemented by effective legal, communications, and financial networks [24,49,95]. The barriers preventing realizations of these systems likely stem from insufficiently complex understandings of, and institutional processes accounting for, the discontinuities separating levels of complexity in language use [51,56,96,97].

To be effective, measurement systems, or better, metasystems, spanning scientific, legal, communications, and financial domains must be capable of functioning across multiple levels of complexity. These metasystems will need to be situated in a larger social context cognizant of needs for the relevant standards bodies, educational programs and textbooks, professional societies, legal and regulatory oversight, accounting and economic applications, etc. Those standards bodies, programs, societies, etc. can be expected to emerge as metrologically traceable implementations of these models gain traction in measuring and managing human [98], social, and natural capital [101,102]. As that happens, it will become increasingly apparent that many of today’s political, regulatory, and business failings are functions of social capital market inefficiencies. When any individual’s, community’s, firm’s, or government’s stock of social capital is routinely measured and traded in public markets, and public opinion gets solidly behind the economics of scientifically tuned and calibrated Genuine Progress Indicators and Happiness Indexes [103], it will be easier to detect and prevent the kinds of failings and abuses economies repeatedly suffer [24,49,104,105].

If our methods of measurement define who we are and what we value, who are we when we willingly continue measuring educational, health, human resource, social service, and environmental quality outcomes in units that vary wildly in quality and size? Who would we be if we were using methods that provide meaningful, constant units with known uncertainties? What do we value when we allow unfair measurements to dictate policies and practices that surely could be improved if higher quality information was available to decision makers? Is this our sense of fair dealing? Or should we redouble our efforts to establish ethical comparisons as our standard background assumptions, taking advantage of recent technical accomplishments and insights into the need for balancing harmony and dissonance, consensus and disagreement, across multiple levels of complexity in our fundamentally symbolic social systems of systems?

References

4. L. Finkelstein, Measurement 42(9), 1270 (2009)


15. K. Sijtsma, Measurement 42, 1209 (2011)


32. M.R. Wilson, Measurement 46, 3766 (2013)


42. D. Andrich, Psychometrika 43(4), 561 (1978)


44. D. Andrich, Medical Care 42(1), 1-7 (2004)

45. D. Andrich, Psychometrika 75(2), 292 (2010)

46. P. Ricoeur, Political and Social Essays (eds Stewart, D. & Bien, J.) (Ohio University Press, Athens, Ohio, 1974)
54. E.B. Andersen, Psychometrika 42(1), 69 (1977)
64. G. Rasch, Probabilistic Models for Some Intelligence and Attainment Tests, (reprint, University of Chicago Press, 1980) (Danmarks Pedagogiske Institut, Copenhagen, Denmark, 1960)
79. E. Hutchins, Mind, Culture, and Activity 19, 314 (2012)
83. O.D. Duncan, Notes on Social Measurement (Russell Sage Foundation, New York, 1984)
91. T. Hancock, Health Promotion International 16(3), 275-80 (2001)
97. A. Woolley, E. Fuchs, Organization Science 22(5), 1359 (2011)