urn:nbn:de:gbv:ilm1-2011imeko-017:2

Joint International IMEKO TC1+ TC7+ TC13 Symposium
August 31st – September 2nd, 2011, Jena, Germany
urn:nbn:de:gbv:ilm1-2011imeko:2

MEASUREMENT, METROLOGY AND
THE COORDINATION OF SOCIOTECHNICAL NETWORKS

William P. Fisher, Jr.[1]

[1]Graduate School of Education, University of California – Berkeley, Berkeley, California, USA

Abstract – Exact, reproducible and well maintained international standards are needed not only in the natural sciences and engineering, but are of increasing concern today in education, health care, government, business intelligence, and the economy at large. Metrological practice gives rise to complex adaptive systems in which meaningful relationships are socially conceived and brought to life. All scientific units should be established with equal rigor, and all measures should be based on consensus standard metrics valued for their accuracy, stability and availability. Fifty years of research and practice employing probabilistic models of psychosocial constructs document individual-level structural invariances capable of supporting metrological traceability. Complex adaptive systems of this kind offer new opportunities for extending the true union of mathematics and measurement into the psychosocial sciences.

Keywords (not more than three): traceability, complex adaptive systems, psychosocial sciences

1. INTRODUCTION

On page 145 of his book, The Mathematics of Measurement: A Critical History, John Roche [1] quotes Charles de La Condamine (1701-1774), who, in 1747, wrote:

> It is quite evident that the diversity of weights and measures of different countries, and frequently in the same province, are a source of embarrassment in commerce, in the study of physics, in history, and even in politics itself; the unknown names of foreign measures, the laziness or difficulty in relating them to our own give rise to confusion in our ideas and leave us in ignorance of facts which could be useful to us.

Roche then explains what de La Condamine is driving at, saying:

> For reasons of international communication and of civic justice, for reasons of stability over time and for accuracy and reliability, the creation of exact, reproducible and well maintained international standards, especially of length and mass, became an increasing concern of the natural philosophers of the seventeenth and eighteenth centuries. This movement, cooperating with a corresponding impulse in governing circles for the reform of weights and measures for the benefit of society and trade, culminated

in late eighteenth century France in the metric system. It established not only an exact, rational and international system of measuring length, area, volume and mass, but introduced a similar standard for temperature within the scientific community. It stimulated a wider concern within science to establish all scientific units with equal rigour, basing them wherever possible on the newly established metric units (and on the older exact units of time and angular measurement), because of their accuracy, stability and international availability. This process gradually brought about a profound change in the notation and interpretation of the mathematical formalism of physics: it brought about, for the first time in the history of the mathematical sciences, a true union of mathematics and measurement.

As it was in the seventeenth and eighteenth centuries for physics, so it has also been in the twentieth and twenty-first centuries for the psychosocial sciences. The creation of exact, reproducible and well maintained international standards is a matter of increasing concern today for the roles they will play in education, health care, the work place, business intelligence, and the economy at large.

The contemporary economic crises are exacerbated by unmet demand for common product definitions, outcome-based budgeting, and interpretable, meaningful measures of impacts in education, health care, social services, environmental management, etc. We need an exact, rational and international system for measuring literacy, numeracy, health, motivations, quality of life, community cohesion, and environmental quality. We need to reinvigorate concern across the sciences to establish all scientific units with equal rigor, and to have all measures based wherever possible on consensus standard metrics valued for their accuracy, stability and availability. We need to replicate in the psychosocial sciences the profound change in the notation and interpretation of the mathematical formalism of physics that occurred in the eighteenth and nineteenth centuries. We need to extend the true union of mathematics and measurement from physics to the psychosocial sciences. The key issue, in short, is the poetic vitality with which things are brought into words in science.

urn:nbn:de:gbv:ilm1-2011imeko-017:2

Joint International IMEKO TC1+ TC7+ TC13 Symposium
August 31st – September 2nd, 2011, Jena, Germany
urn:nbn:de:gbv:ilm1-2011imeko:2

2. A COMPLEX ADAPTIVE LOGIC OF LIVING MEANING

In his history of science, Roche [1] refers to the systematic creation and implementation of universally uniform quantitative units, mathematical notation, and practical applications as "a true union of mathematics and measurement." Mathematically rigorous and universally uniform measurement is valuable for the way it functions as a common currency or *lingua franca*. When meaning can be shared at a social level, behaviours and decisions are coordinated with no need for communicating or negotiating the details of each particular situation. A kind of choral effect is thereby achieved in the sense that harmonious social, political, and economic relationships are produced as many independent actors respond in unison to variations in phenomena, such as time, agricultural production, price fluctuations, or product quality.

Thus it happens that the infrastructure provided by the common language facilitates the spontaneous self-organization of networks of agents and actors capable of transparently coordinated thought and action [2-7]. These networks are increasingly recognized as being of central importance in the process by which phenomena are made to seem naturally universal. Widely adopted and easily available technical standards are created and implemented within and by social networks that habituate themselves to the intuitive application of common languages. The resources and effort put into creating the appearance of natural universality are then selectively ignored in accord with value systems and metaphysical assumptions that prioritize a kind of concrete operational conception of objectivity over a formal or abstract conception.

A number of seemingly insurmountable difficulties are encountered when these problems are taken up in a purely philosophical examination. An alternative approach focuses on the material practices and communications of technicians and scientists as they converse, write memos, give conference presentations, publish papers, secure funding, manage organizations, etc. in the course of making discoveries and bringing them to market in practical applications. Social studies of science in this tradition find science and theory so dependent on the technical means of routinely producing and reproducing phenomena that science is seen to be impossible without the technology and the social networks sustaining it.

Hence, to remind ourselves of these dependencies, the terms technoscience and sociotechnical have come into use. The conception, gestation, birthing, and nurturing of complex adaptive systems constitute a reproductive logic for sociocultural traditions' capacity for collective cognition. Scientific and economic traditions, in particular, form mature self-identities via a mutually implied subject-object relation absorbed into the flow of a dialectical give and take.

2.1. Complex adaptive functionality as a developmental model

By means of this dialectic, complex adaptive systems establish the reproductive viability of their offspring and the coherence of an ecological web of meaningful relationships. There are five moments in the formation and operation of complex adaptive systems, which must be able (1) to identify regularities and patterns in the flow of matter, energy, and information (MEI) in the environment (social, economic, natural, etc.); (2) to produce condensed schematic representations of these regularities so they can be identified as the same if they are repeated; (3) to form reproductively interchangeable variants of these representations; (4) to succeed reproductively by means of the accuracy and reliability of the representations' predictions of regularities in the MEI data flow; and (5) adaptively modify and reorganize representations by means of informational feedback from the environment [8]. All living systems, from bacteria and viruses to plants and animals to languages and economies, are complex adaptive systems characterized by these five features.

In the history of science, technologically-embodied measurement facilitates complex adaptive systems of various kinds. That history can be used as a basis for a meta-theoretical perspective on what measurement must look like in the social and human sciences. Each of the five moments in the formation and operation of complex adaptive systems describes a capacity of measurement systems, in that (1) data flow regularities are captured in initial, provisional instrument calibrations; (2) condensed local schematic representations are formed when an instrument's calibrations are anchored at repeatedly observed, invariant values; (3) interchangeable nonlocal versions of these invariances are created by means of instrument equating, item banking, metrological networks, and selective, tailored, adaptive instrument administration; (4) measures read off inaccurate and unreliable instruments will not support successful reproduction of the data flow regularity, but accurate and reliable instruments calibrated in a shared common unit provide a reference standard metric that enhances communication and reproduces the common voices and shared identities of communities of research and practice; and (5) consistently inconsistent anomalous observations provide feedback on new possibilities for as yet unrecognized data flow regularities that might be captured in new constructs' calibrations.

Measurement in the social sciences is in the process of extending this functionality into practical applications in business, education, health care, government, and elsewhere. Over the course of the last 50 years, measurement research and practice has already iterated many times through these five moments. In the coming years, a new critical mass will be reached in this process, systematically bringing about scale-of-magnitude improvements in the efficiency of intangible assets markets.

urn:nbn:de:gbv:ilm1-2011imeko-017:2

Joint International IMEKO TC1+ TC7+ TC13 Symposium
August 31st – September 2nd, 2011, Jena, Germany
urn:nbn:de:gbv:ilm1-2011imeko:2

How? What does a "data flow regularity" look like? How is it condensed into a a schematic and used to calibrate an instrument? How are local schematics combined together in a pattern used to recognize new instances of themselves? More specifically, how might enterprise resource planning (ERP) software (such as SAP, Oracle, or PeopleSoft) simultaneously provide both the structure needed to support meaningful comparisons and the flexibility needed for good fit with the dynamic complexity of adaptive and generative self-organizing systems?

Prior work in this area proposes a dual-core, loosely coupled organization using ERP software to build social and intellectual capital, instead of using it solely as an IT solution addressing organizational inefficiencies [9]. The adaptive and generative functionality [10] provided by probabilistic measurement models [11-15] makes it possible to measure and manage intra- and inter-organizational interoperability [16] at the same time that social and intellectual capital resources are augmented.

First, the data flow regularities identified and then captured in schematic representations are illustrated graphically. Probabilistic measurement models are usually presented mathematically in the form of equations and the statistics resulting from their application to data. An alternative form of mathematical presentation offers images of the patterns expected and diagnosed in the application of the model. After explicating this approach to understanding the modelled structural invariances, attention will turn to the means by which practical implementations of interchangeable nonlocal versions of these invariances may augment social and intellectual capital resources.

2.2. Graphical schema

Fig. 1 shows a data matrix of 24 columns by 40 rows. The individual observations are responses in one of three categories to questions (the columns) on a survey given by participants in a research study (the rows). These data are a small subset extracted randomly (without replacement) from a much larger data set involving hundreds of items and thousands of survey respondents. The larger data set was analyzed using Rasch measurement software [17-19] to calibrate an instrument measuring the efforts special education personnel make to partner with the parents of the children with disabilities enrolled in their programs [20]. Overall model fit was satisfactory, with measurement reliabilities ranging from 0.95 to 0.97. The figure illustrates the regularity of the pattern in the data captured in the flow of providing special education services. Blank spaces indicate data missing due to the administration of separate forms and to skipped items.

The individual responses shown in Fig. 1 are ordered according to the individual person-by-item response probabilities determined from the larger complete data set. The person with the highest measure

and the highest probability of responding in the highest category (3) across all of the items is at the top. The person with the lowest measure and the lowest probability of responding in the highest category on any item is at the bottom. The survey item that is easiest to agree with and is most likely to provoke responses of 3 is in the far left column, and the one that is hardest to agree with is on the far right.

To illustrate the consistency of the stochastic pattern in the data, responses of 3 are coloured red, responses of 2 are black, and responses of 1 are blue. This pattern may not look like much to those unfamiliar with probabilistic measurement theory, its capacity to take missing data into account, and the role of error ranges and confidence intervals in interpreting measures. It may seem as though this pattern is an accident that occurred once, never to appear again. Less radic ally, there may be room for hypothesizing the existence of more than one construct in these data, though extensive evaluations of the data so far tend to rule out that possibility. As Taylor warns, the system has to be able to avoid the errors of mistaking randomness for order, and vice versa.

Reading down from the top of Fig. 1, 3s tend to drop to 2s, and 2s to 1s, from the right, with the 2s and 1s progressing further and further to the left from the right the further down the page the cases are. Being drawn from a much larger data set (with dozens more columns and thousands more rows), this subsample is only roughly representative of the general pattern, as
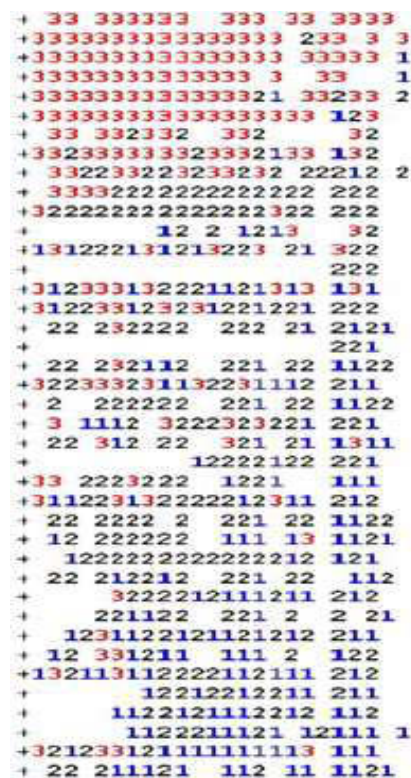

Fig. 1. First example of an ordered matrix of three-category data

urn:nbn:de:gbv:ilm1-2011imeko-017:2

Joint International IMEKO TC1+ TC7+ TC13 Symposium
August 31st – September 2nd, 2011, Jena, Germany
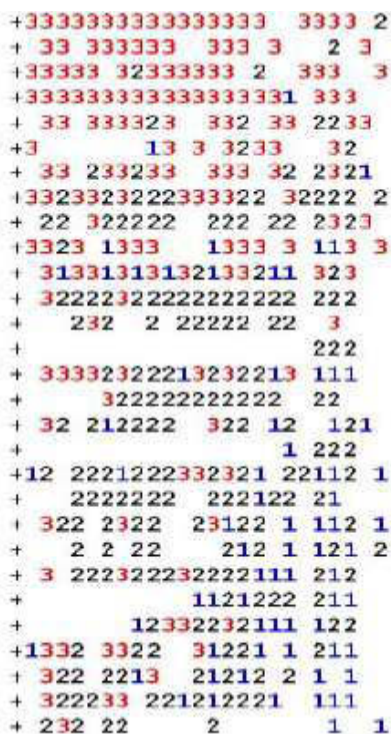urn:nbn:de:gbv:ilm1-2011imeko:2

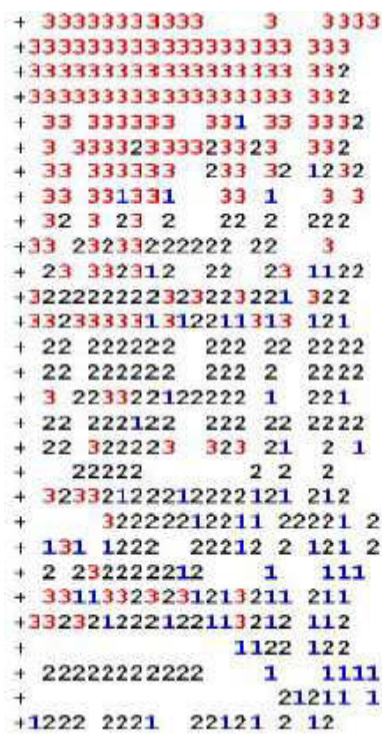Fig. 2. Second example of an ordered matrix of three-category data

Fig. 3. Third example of an ordered matrix of three-category data

will be shown in greater detail. For the moment, it suffices to note that most cases have response patterns that accord with the overall pattern, and so do not exhibit left-to-right or top-to-bottom progressions from 1s to 3s.

And so the question arises: to what extent is this pattern repeated across other samples from the larger data set? Figs. 2 and 3 show the same items in the same order, but with entirely different random samples of respondents, as can be seen from the different entry order sequence numbers at the far left in each figure.

The repetition of the same pattern across three different samples of survey respondents is evidence supporting the inference that the regularity identified in the first data set has been reproduced twice more. Any single person's responses might seem completely random, but repeated posing of the questions to even a small sample of 40 gave rise to the consistent pattern found in Figs. 1, 2 and 3. These redundancies in the flow of experience in this domain are then seen to condense into a general schematic with their repetition in two more samples of respondents. Plainly, the order is not random, and the randomness within the order is evaluated in terms of reliability and model fit in routine analyses implementing Rasch models.

The same thing happens when the rows are held constant, and the columns are varied. In other words, the constancy of the pattern observed when different people answer the same questions is also found when the same people answer different questions designed to measure the same construct. And so, it also happens that the same patterns are also reproduced across dif-

ferent samples of survey respondents answering entirely different sets of questions (all of which are carefully composed, piloted, and calibrated into a bank of items repeatedly demonstrated as measuring the same thing).

These images convey an admittedly impressionistic image of the point that there is a real construct represented by these data, one that asserts itself as an independent entity separate from the local contingencies of its origin. Now the question is, to what extent do these rough schematic patterns support the inference that different sets of items can measure the same thing across different samples of people in the same unit?

2.3 From condensed local schema to interchangeable nonlocal ones

Taking up Taylor's [8] language again, the regularity in the flux of data identified in Figs. 1-3 suggests the form of a schema that will enable the system to recognize the pattern if it is encountered again. The second moment in complex adaptive functionality is obtained when the data are compressed via an algorithm that enables the system to recognize the pattern again, and unfold its details efficiently. The more reducible the data are, the greater the efficiency obtained in the system's functioning. Compression is arguably most efficient when its data reduction algorithm is able to make use of sufficient statistics. Rasch models, for which raw scores are minimally sufficient statistics (both necessary and sufficient), enable in-

urn:nbn:de:gbv:ilm1-2011imeko-017:2

Joint International IMEKO TC1+ TC7+ TC13 Symposium
August 31ˢᵗ – September 2ⁿᵈ, 2011, Jena, Germany
urn:nbn:de:gbv:ilm1-2011imeko:2

formation processing that filters noise, and creates order from chaos.

At this point in his description of the five features of complex adaptive systems, Taylor [8] notes that "the schemata or algorithms in complex adaptive systems are emergent and can change," and so, he suggests, they cannot be preprogrammed or fixed. But adaptive, emergent, and changeable algorithms can, and routinely are, preprogrammed and fixed in computerized adaptive testing and in applications of predictive construct theories, as is suggested by the nature of the patterns shown in Figs. 1-3.

There are two ways of accomplishing this preprogramming [10], with one less efficient than the other. In the less efficient option, the one most often used in computerized adaptive testing, the preprogramming takes the form of experimentally-determined item calibrations. In this context, all items must be calibrated on actual data before they can be used, but once calibrated, comparability no longer requires that the same specific items be employed in measuring any one person.

The second, more efficient option focuses on understanding the construct measured well enough to specify the item component traits that make items easy or difficult, agreeable or disagreeable. With this information in hand, items can be written and administered from theory, with quality checks determined by whether the intended response probabilities are consistently obtained or not.

Furthermore, unidimensional schema will themselves form consistent and repeating patterns relative to each other, giving rise to multidimensional models and associated schema at higher levels of complexity [21].

Whether in the item banking or the item specification contexts, or relative to unidimensional or multidimensional models, as Taylor [8] recognizes, "For a schema to be functional, relevant data have to be compressed as much as possible." In this compression, Taylor points out, the schema must process information in a way that creates order from noise, that is adaptive in the sense of allowing for change, and that is efficient enough to maximize the available storage space and so extend memory. The phenomena represented by structurally invariant data fitting probabilistic measurement models are forms of stochastic resonance creating order from noise [22, 23], adaptable to changing circumstances, and, via the principle of statistical sufficiency, reducing data volume dramatically with no loss of information [24-28].

3. MODELING THE SCHEMA

A kind of wilful ignorance or pathological state of denial seems to prevent many in research and business intelligence from perceiving and acting on invariant patterns like those in Figs. 1-3. At the very least, if more researchers were aware that test, survey, and assessment data can organize themselves like this,

more organizations would be deliberately constructing and capitalizing on these patterns in the name of quality improvements and research advances in communications, education, health care reform, and business intelligence. But few are looking for these patterns, and those who are looking for them do not seem to realize their importance, or do not know how to tell their story. Finding ways to tell the story is a major challenge of our time.

The coherence with which things speak to us across data sets does not, of course, happen by itself. Questions can be framed to support, to be indifferent to, or to defeat, the purpose of measurement. To measure efficiently, precisely, and with the advantages attending predictive control of the construct, the composition and phrasing of questions has to be informed by a theory of what counts as less and more of the thing to be measured.

The preparation of a medium in which a self-organizing conceptual entity like a measured construct can manifest itself is akin to laboratory work in which a micro-environment for a form of life is created. What we do, in effect, is pose questions to people likely to be caught up in the play of the intended construct's language game. In the back and forth of the question and answer process, everyone gets caught up in the construct's self-representative activity. The thing itself then is able to write its story on the abstract conceptual medium provided to it, acting through the participants in the dialogue.

The point of calibrating instruments using Rasch models is to test the hypothesis that this kind of regularity exists and will repeat itself in patterns of data across samples of persons and items. These invariances of the persons over the items and the items over the persons are what Rasch [11, 29] was referring to when, referring to a model for measuring reading ability, he wrote

> On the basis of [one of the equations in the model] we may estimate the item parameters independently of the personal parameters, the latter having been replaced by something observable, namely, by the individual total number of correct answers. Furthermore, on the basis of [the next equation] we may estimate the personal parameters without knowing the item parameters which have been replaced by the total number of correct answers per item. Finally, [the third equation] allows for checks on the model [another equation] which are independent of all the parameters, relying only on the observations.

Figs. 1-3 illustrate Rasch's first sentence here, in that they show that the same data pattern forming across samples for a group of items. Because the pattern stays the same across the observed responses, we can estimate the calibration parameters with no concern for the particular persons measured. Then, copying the

urn:nbn:de:gbv:ilm1-2011imeko-017:2

Joint International IMEKO TC1+ TC7+ TC13 Symposium
August 31st – September 2nd, 2011, Jena, Germany
urn:nbn:de:gbv:ilm1-2011imeko:2

entire data file of 2,500 survey respondents and 78 items three times, and then removing different rows and columns from each of them produces three completely different subsamples of cases and items, as shown in Fig. 4.

This figure conveys Rasch's point that a real construct asserts itself as an independent entity that separates from the local contingencies of its origin. Different sets of items can measure the same thing across different samples of people in the same unit, just as different samples of people can calibrate different sets of items to measure in the same unit. The pattern in the data stays the same across groups of items for a given sample of people, and vice versa. Accordingly, we can estimate measures without knowing which items are involved, and items can be calibrated independent of the particular persons sampled. Finally, as is illustrated in Fig. 5, we can evaluate the clarity of our thinking, and the model, by examining the observ-
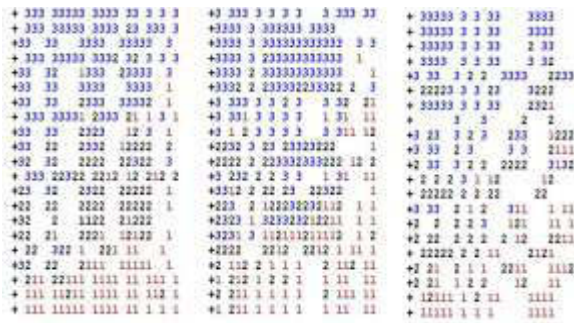

Fig. 4. Three ordered matrices: Different persons, different items, same construct

ations alone, apart from any concern with the particulars of who answered the questions or which questions were asked.

Fig. 5 illustrates the overall general pattern. Holding the items and/or the calibrations in the columns constant, as measures on the vertical axis go up, more and more of the rows become red. Holding the persons and/or the measures in the rows constant, more and more of the columns are blue as calibrations on the horizontal axis move further to the right and become more difficult.
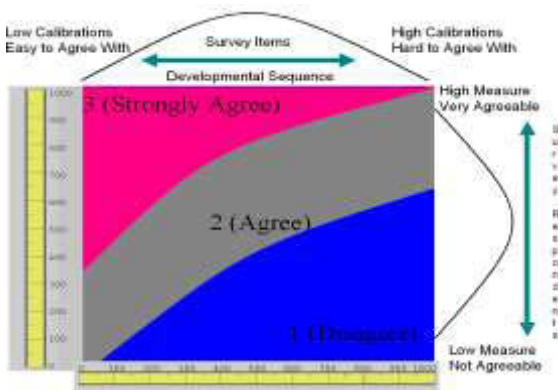

Fig. 5. The general pattern, annotated and scaled

Furthermore, the calibrations do not simply stay in the same order across samples, they retain the same unit size; i.e., they are invariantly positioned relative to one another, within the range of error, and are not just invariantly ordered. Fig. 6 shows the scatter plot of the Sample 1 calibrations vs the Bank Calibration Anchors. The confidence intervals are fairly wide, as is expected given the error and reliabilities obtained with Sample 1's per-item sample sizes of about 40. In accord with reliability theory, the confidence intervals become progressively narrower as sample size and reliability increase [30].

The two items falling outside the confidence intervals had only 3-12 responses in all three samples, where all but one of the other 22 items had at least 30 in all three samples. Despite having only about 40 cases per item, on average, these 22 of the 24 item calibrations correlate 0.81 to 0.90 with each other

TABLE I. Correlations of Fig. 1-3 independent sample calibrations

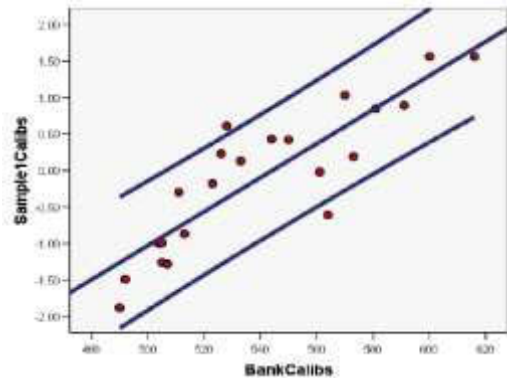|	Subsample 2	Subsample 3	Total Sample
Subsample 1	.817	.902	.875
Subsample 2		.812	.908
Subsample 3			.895

Fig. 6. Scatter plot of the .88 correlation

across samples; they correlate 0.88 to 0.91 with the established bank anchors for these items (see Table I). When disattenuated for error [31-33], all of these correlations are 1.00.

Figs. 4-6 illustrate the fourth and fifth moments in the formation and operation of complex adaptive systems. The constancy of the ideal abstract pattern across samples of people and items provides the reliability evidence and a basis for predictive construct theory needed to support the calibration of a reference standard metric representing a stable and regular flow of information. The system is not dependent on items of a particular content, and so it can adapt to changingg circumstances as those measured grow and develop. Fifth, sensitivity to anomalous inconsistencies, such as the outlier points in Fig. 6, set the stage for reconsiderations of the data pattern supporting infer-

urn:nbn:de:gbv:ilm1-2011imeko-017:2

Joint International IMEKO TC1+ TC7+ TC13 Symposium
August 31ˢᵗ – September 2ⁿᵈ, 2011, Jena, Germany
urn:nbn:de:gbv:ilm1-2011imeko:2

ences relative to the given construct, opening the door to critical re-evaluations of both individual-level and group-level interpretations of the measured amounts.

4. CONCLUSION

The invariances of item calibrations across samples, and of measures across item samples, and of both item and person estimates across samples of judges rating performances, have supported the emergence of a testing industry that capitalizes on the repeatable confirmation of predictable data patterns. Study of these kinds of patterns leads to opportunities for

- Mass customization
  - Items individually tailored to respondents' needs
  - All measures expressed in single common metric
- Data quality assessment
  - Internal consistency evaluation
  - Goes beyond checking codes
- Data volume reduction
  - One measure per person, unit, division, or facility
  - Instead of multiple numbers for each, per item
- Meaningful measures
  - A substantial thing measured that adds up the way the numbers do
  - Not numbers that vary in meaning relative to irrelevant factors
- Fairer measures
  - Data from judges rating performances can be evaluated for internal consistency and leniency/severity estimated
  - Individual measures can be adjusted across raters, creating level playing field

But beyond these practical applications, what is it that makes question and answer interactions of this kind exhibit this kind of invariance? The reproducibility of this kind of pattern across samples of people and items belonging to the same populations demands an explanation. Is there any identifiable common thread that unfolds in a predictable way as the item calibrations increase?

Items giving expression to measured constructs do not take random orders that nonetheless stay constant across samples. What sense would there be in a consistency of the examinee or respondent order across samples of items that was just a function of abilities, behaviours, or attitudes that remain at a constant level no matter what questions are asked? Most usually, some items and people can be seen to possess more and less of some one thing to which all of them are interrelated. The question to ask is then, what is there about the questions associated with these patterns that allows them to function as media for the repeatedly iterable self-representative, self-organizing, adaptive play of measured constructs?

Sociohistorical descriptions of successful metrological programs need to be translated into a practical program for psychosocial metrology. In his study of the emergence of electrical metrology in the late 19ᵗʰ century, Schaffer [4] remarks that

> The physical values which the laboratory fixes are sustained by the social values which the laboratory inculcates. Metrology has not often been granted much historical significance. But in milieux such as those of Victorian Britain the propagation of standards and values was the means through which physicists reckoned they could link their work with technical and economic projects elsewhere in their society. Instrumental ensembles let these workers embody the values which mattered to their culture in their laboratory routines. Intellectualist condescension distracts our attention from these everyday practices, from their technical staff, and from the work which makes results count outside laboratory walls.

Metrology is being granted significantly more historical significance in the context of the social studies of science that it had received previously. Shapin [5], for instance, documents the way that the technical means by which laboratory phenomena are produced have often been literally cut out of the picture. Recognition of the vital role played by metrology is a significant development with major implications for the emergence of improved measurement in education, health care, psychology, and the social sciences.

Though few are yet aware of the converging trends, repeated application of Rasch models to data from different test, survey, and assessment instruments measuring the same things in the same populations must inevitably lead to increased recognition of the potential for and value of uniform metric standards [34-37]. When this happens, previous historical processes will be brought to bear in the human, social, and environmental sciences. For instance, writing in 1881 on the converging results in electrical measurement, and echoing de la Condamine's more broadly framed remarks from the previous century, Everett [38] observed that

> the existence of quantitative correlations between the various forms of energy, imposes upon men of science the duty of bringing all kinds of physical quantity to one common scale of comparison.

Quantitative correlations between the various forms of reading ability and text difficulty have already imposed on some a sense of a duty and an opportunity for bringing all kinds of language comprehension

urn:nbn:de:gbv:ilm1-2011imeko-017:2

Joint International IMEKO TC1+ TC7+ TC13 Symposium
August 31st – September 2nd, 2011, Jena, Germany
urn:nbn:de:gbv:ilm1-2011imeko:2

quantity to a single common scale of comparison [39]. Similarly, quantitative correlations between various approaches to human development and maturation have led to similar instrument equatings [40, 41].

And just as was the case in Victorian Britain, the propagation of standards and values is a means by which researchers in the human and social sciences will be able to link their work with technical and economic projects elsewhere in society. The numeric values assigned to quantitative amounts in systematically developed and maintained units of measurement are not devoid of aesthetic, moral, economic, and political implications, however much we typically assume that they are. Ongoing studies in the co-production of order across domains, and especially in the economic implications of metrology [42-49], must inevitably reveal the ways in which values of many kinds are embodied in the results of laboratory routines exported into the world.

Marked differences in the quality of values will be found in contrasts of results expressed in purely local and ordinal terms with those expressed in general and interval metrics. In the former cases, the Socratic midwife must declare a stillbirth. The failure to define a unit that transcends the local particulars of who, what, and when means that the results, however valuable they might be, will not be able to move into the world as readily interpretable, meaningful significations taking on a life of their own. In the latter cases, those in which results are expressed in a unit of measurement that remains constant across instruments and samples (within a range of uncertainty), meaning will be found well-formed, with everything it needs to sustain itself outside of the laboratory in which it originated. Technology road maps have proven their value as guides for coordinating expectations and aligning resources across firms in a number of industries [49]; the trajectory defined by measurement as a complex adaptive system opens the door to other similar kinds of road maps for education, health care, government, etc.

This kind of a metasystem for complex adaptive functionality in research is implied in what Roche calls a "true union of mathematics and measurement" [1]. Complex adaptive systems are effectively constituted of such a union. Probabilistic conjoint models for fundamental measurement are poised to extend this functionality into the human, social, and environmental sciences. A renewal of the true union of mathematics and measurement, of quality and quantity, and of the Socratic midwife and the Socratic gadfly, may yet lead to a new cultural, scientific, and economic renaissance.

ACKNOWLEDGEMENT

REFERENCES

[1]   J. Roche, The mathematics of measurement: A critical history. The Athlone Press, London, 1998.

[2]   B. Latour, Reassembling the social: An introduction to actor-network-theory, Oxford University Press, Oxford, 2005.

[3]   D. Ihde, E. Selinger (Eds.), *Chasing technoscience: Matrix for materiality*. (Indiana Series in Philosophy of Technology). Bloomington, Indiana: Indiana University Press, 2003.

[4]   S. Schaffer, "Late Victorian metrology and its instrumentation: A manufactory of Ohms", In R. Bud & S. E. Cozzens (Eds.), *Invisible connections: Instruments, institutions, and science* (pp. 23-56). Bellingham, Washington, USA, SPIE Optical Engineering Press, 1992.

[5]   S. Shapin, "The invisible technician", *American Scientist*, vol. 77, pp. 554-563, 1989.

[6]   E. Mendelsohn, "The social locus of scientific instruments", In R. Bud & S. E. Cozzens (Eds.), *Invisible connections: Instruments, institutions, and science* (pp. 5-22). Bellingham, Washington, USA, SPIE Optical Engineering Press, 1992.

[7]   J. O'Connell, "Metrology: The creation of universality by the circulation of particulars", *Social Studies of Science*, vol. 23, pp. 129-173, 1993.

[8]   M.C. Taylor, *The moment of complexity: Emerging network culture.* University of Chicago Press, Chicago, 2003.

[9]   C.A. Lengnick-Hall, M.L. Lengnick-Hall, S. Abdinnour-Helm, "The role of social and intellectual capital in achieving competitive advantage through enterprise resource planning (ERP) systems", *Journal of Engineering Technology Management*, vol. 21, pp. 307-330, 2004.

[10] A.J. Stenner, M. Stone, "Item specification vs. item banking", *Rasch Measurement Transactions*, vol. 17, no. 3, pp. 929-30, 2003.

[11] G. Rasch, *Probabilistic models for some intelligence and attainment tests* (Reprint, with Foreword and Afterword by B. D. Wright, Chicago: University of Chicago Press, 1980). Copenhagen, Denmark: Danmarks Paedogogiske Institut, 1960.

[12] D. Andrich, *Rasch models for measurement.* (Vols. series no. 07-068). Sage University Paper Series on Quantitative Applications in the Social Sciences. Beverly Hills, California: Sage Publications, 1988.

[13] T. Bond, C. Fox, *Applying the Rasch model: Fundamental measurement in the human sciences, 2d edition*. Mahwah, New Jersey: Lawrence Erlbaum Associates, 2007.

[14] B.D. Wright, "Additivity in psychological measurement", In E. Roskam (Ed.), *Measurement and personality assessment* (pp. 101-112). North Holland: Elsevier Science Ltd, 1985.

[15] B.D. Wright, "Fundamental measurement for psychology", In S. E. Embretson & S. L. Hershberger (Eds.), *The new rules of measurement: What every educator and psychologist should know* (pp. 65-104), Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1999.

[16] G. Weichhart, T. Feiner, C. Stary, "Implementing organisational interoperability--The SUddEN approach", *Computers in Industry*, vol. 61, pp. 152-160, 2010.

urn:nbn:de:gbv:ilm1-2011imeko-017:2

Joint International IMEKO TC1+ TC7+ TC13 Symposium
August 31st – September 2nd, 2011, Jena, Germany
urn:nbn:de:gbv:ilm1-2011imeko:2

[17]   J.M. Linacre, *A user's guide to WINSTEPS Rasch-Model computer program, v. 3.72.0.* Chicago, Illinois: Winsteps.com, 2011.

[18]   M.L. Wu, R. Adams, M.R. Wilson, S.A. Haldane, *ACER ConQuest Version 2: Generalised item response modelling software* [computer program]. Camberwell: Australian Council for Educational Research, 2007.

[19]   D. Andrich and B. Sheridan, *RUMM 2030*, RUMM Laboratory Pty Ltd: Perth, WA, 1997-2011.

[20]   B. Elbaum, W.P. Fisher, Jr., W.A. Coulter, W. A. (2011). "Measuring schools' efforts to partner with parents of children served under IDEA", *Journal of Applied Measurement*, in review, 2011.

[21]   D. Briggs, M. Wilson, "Understanding Rasch measurement: An introduction to multidimensional measurement using Rasch models", *Journal of Applied Measurement*, vol. 4, no. 1, pp. 87-100, 2003.

[22]   W.P. Fisher, Jr., "Stochastic resonance and Rasch measurement", *Rasch Measurement Transactions*, vol. 5, no. 4, pp. 186-187, 1992.

[23]   W.P. Fisher, Jr., "Stochastic and historical resonances of the unit in physics and psychometrics", *Measurement: Interdisciplinary Research & Perspectives*, vol. 9, pp. 46-50, 2011.

[24]   E. B. Andersen, "Sufficient statistics and latent trait models", *Psychometrika*, vol. 42, no. 1, pp. 69-81, 1977.

[25]   E. B. Andersen, "Sufficient statistics in educational measurement", In G. N. Masters & J. P. Keeves (Eds.), *Advances in measurement in educational research and assessment* (pp. 122-5). New York, Pergamon, 1999.

[26]   G.H. Fischer, "On the existence and uniqueness of maximum-likelihood estimates in the Rasch model", *Psychometrika*, vol. 46, no. 1, pp. 59-77, 1981.

[27]   G.H. Fischer, "The derivation of polytomous Rasch models", In G. H. Fischer & I. Molenaar (Eds.), *Rasch models: Foundations, recent developments, and applications* (pp. 293-305). New York, Springer-Verlag, 1995.

[28]   B.D. Wright "Rasch model from counting right answers: Raw scores as sufficient statistics", *Rasch Measurement Transactions*, vol. 3, no.2, p. 62, 1989.

[29]   G. Rasch, "On general laws and the meaning of measurement in psychology", In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability* (pp. 321-333). Berkeley, California: University of California Press, 1961.

[30]   W.P. Fisher, Jr., B. Elbaum, W. A. Coulter, "Reliability, precision, and measurement in the context of data from ability tests, surveys, and assessments", *Journal of Physics, Conference Series,* vol. 238, no. 1, http://iopscience.iop.org/1742-6596/238/1/012036/pdf/1742-6596_238_1_012036.pdf, 2010.

[31]   P.M. Muchinsky, "The correction for attenuation", *Educational and Psychological Measurement*, vol. 56, no. 1, pp. 63-75, 1996.

[32]   R.E. Schumacker, "Disattenuating correlation coefficients", *Rasch Measurement Transactions*, vol. 10, no. 1, p. 479, 1996.

[33]   B.D. Wright "Errors, variances and correlations", *Rasch Measurement Transactions*, vol. 5, no. 2, p. 147, 1991

[34]   W.P. Fisher, Jr., "Physical disability construct convergence across instruments: Towards a universal metric", *Journal of Outcome Measurement*, vol. 1, no. 2, pp. 87-113, 1997.

[35]   W.P. Fisher, Jr., "What scale-free measurement means to health outcomes research", *Physical Medicine & Rehabilitation State of the Art Reviews*, vol. 11, no. 2, pp. 357-373, 1997.

[36]   W.P. Fisher, Jr., "Foundations for health status metrology: The stability of MOS SF-36 PF-10 calibrations across samples", *Journal of the Louisiana State Medical Society*, vol. 151, no. 11, pp. 566-578, 1999.

[37]   W.P. Fisher, Jr., "Invariance and traceability for measures of human, social, and natural capital: Theory and application", *Measurement*, vol. 42, no. 9, pp. 1278-1287, 2009.

[38]   C. Smith, M. N. Wise, *Energy and empire: A biographical study of Lord Kelvin*. Cambridge: Cambridge University Press, 1989.

[39]   A.J. Stenner, H. Burdick, E.E. Sanford, D.S. Burdick, "How accurate are Lexile text measures?", *Journal of Applied Measurement*, vol. 7, no. 3, pp. 307-322, 2006.

[40]   T.L. Dawson, "Assessing intellectual development: Three approaches, one sequence", *Journal of Adult Development*, vol. 11, no. 2, pp. 71-85, 2004.

[41]   T.L. Dawson, "A stage is a stage is a stage: A direct comparison of two scoring systems", *Journal of Genetic Psychology*, vol. 164, pp. 335-364, 2003.

[42]   K. Alder, *The measure of all things: The seven-year odyssey and hidden error that transformed the world.* New York, The Free Press, 2002.

[43]   H. Harbers, *Inside the politics of technology: Agency and normativity in the co-production of technology and society*. Amsterdam, Netherlands, Amsterdam University Press, 2005.

[44]   S. Jasanoff, *States of knowledge: The co-production of science and social order.* (International Library of Sociology). New York, Routledge, 2004.

[45]   S. Jasanoff, *Designs on nature: Science and democracy in Europe and the United States*. Princeton, New Jersey, Princeton University Press, 2005.

[46]   W. J. Ashworth, "Metrology and the state: Science, revenue, and commerce", *Science*, vol. 306, no. 5700, pp. 1314-1317, 2004.

[47]   G.J. Beges, Drnovsek, L.R. Pendrill, "Optimising calibration and measurement capabilities in terms of economics in conformity assessment", *Accreditation and Quality Assurance: Journal for Quality, Comparability and Reliability in Chemical Measurement*, vol. 15, no.3, pp. 147-154, 2011.

[48]   J.A. Birch, Benefit of legal metrology for the economy and society (International Committee of Legal Metrology). Hunters Hill, Australia: International Organization of Legal Metrology (accessed 28 Oct 2008   No. http://www.oiml.org/publications/E/birch/E002-e03.pdf), 2003.

[49]   P. Miller, T. O'Leary, "Mediating instruments and making markets: Capital budgeting, science and the economy", *Accounting, Organizations, and Society*, vol. 32, no. 7-8, pp. 701-34, 2007.

Author(s): William P. Fisher, Jr., Berkeley Evaluation and Assessment Research Center, Graduate School of Education, University of California – Berkeley, (94720, Berkeley, California, USA) +1-919-599-7245, wfisher@berkeley.edu.