NÚMERO ESPECIAL | 2019
ISSN 2183-0940

REVISTA TMQ

TECHNIQUES, METHODOLOGIES AND QUALITY

SPECIAL ISSUE
Metrology on Health

INVITED EDITOR
MARIA DO CÉU FERREIRA
INSTITUTO PORTUGUÊS DA QUALIDADE

PUBLISHERS
ANTÓNIO RAMOS PIRES
INSTITUTO POLITÉCNICO DE SETÚBAL

MARGARIDA SARAIVA
UNIVERSIDADE DE ÉVORA

ÁLVARO ROSA
ISCTE-IUL

REDE DE INVESTIGADORES DA QUALIDADE

A R**EVISTA TMQ** - TECHNIQUES METHODOLOGIES AND QUALITY, é uma marca registada no **INPI** com o Nº ***614089***.
A Revista TMQ sujeita os artigos publicados a dupla revisão por pares, estando disponível online em: https://publicacoes.riqual.org

A TMQ está indexada a Latindex - Sistema regional de informações on-line para revistas de investigação científica nos países da América Latina, Caribe, Espanha e Portugal.

The REVISTA TMQ TECHNIQUES METHODOLOGIES AND QUALITY is a registered trademark with **INPI** under Nº ***614089***.
The TMQ is a peer-reviewed and publicly available journal, being available online at: https://publicacoes.riqual.org

TMQ is indexed at Latindex - Regional system of online information for scientific research journals in the countries of Latin America, the Caribbean, Spain and Portugal.

FICHA TÉCNICA:
Título: Número Especial - Metrologia na Saúde| 2019
ISSN: 2183-0940
Editora: RIQUAL - Rede de Investigadores da Qualidade
Paginação e produção gráfica: APQ
e-mail: info@riqual.org

TECHNICAL DATA SHEET:
Title: Special Issue – Metrology on Health | 2019
ISSN: 2183-0940
Publisher: RIQUAL - Network of Quality Researchers
Pagination and graphic production: APQ
e-mail: info@riqual.org

TMQ – TECHNIQUES, METHODOLOGIES AND QUALITY

TMQ, NÚMERO ESPECIAL, METROLOGIA NA SAÚDE, 2019 **ISSN: 2183-0940**

AUTORES

Baki Karaböces - TÜBİTAK UME - National Metrology Institute

Elisabeth Costa Monteiro - Federal University of Rio de Janeiro

Elsa Batista - Instituto Português da Qualidade

J. Melin - Rise Research Institutes of Sweden

Karen Tee - Foundry

LR Pendrill - RISE- Research Institutes of Sweden

Maria do Céu Ferreira - Instituto Português da Qualidade

Skye Barbic - University of British Columbia

Steeve Mathias - University of British Columbia

Stefan J. Cano - British Psychological Society

William P. Fisher - University of California, Berkeley

Índice

Editorial 7

Patient-centered outcome measurement in psychiatry: How metrology can optimize health services and outcomes 10

Skye Barbic || Stefan J. Cano || Karen Tee || Steve Mathias

1. Introduction 11
2. Engaging patient partners in outcome conceptualization and prioritization in mental health 11
3. The application of modern measurement methods to develop, test, and use patient centered outcome measures in clinical practice 12
4. Patient-centered outcome service delivery and funding allocation 13
5. Example of designing mental health systems with the principles of metrology: Foundry 14
   5.1 Toolbox
      5.1.1. Hypothesis-driven, patient-centered outcome measurement 14
      5.1.2. Clinically meaningful application of patient-centered outcome measurement 17
6. Conclusion 17

References 18

Authors Profiles 19

A Nondualist Social Ethic: Fusing Object and Subject Horizons in Measurement 21

William P. Fisher

1. Dualism and Nondualism 22
2. From individual experience to institutions and systems 26
3. Conclusion 31

References 34

Author Profiles 40

Magnetic Quantities: Healthcare Sector Measuring Demands and International Infrastructure for providing Metrological Traceability 42

Elisabeth Costa Monteiro

1. Introduction 43
2. Health technologies based on magnetic quantities demanding metrological traceability 43
   2.1 Biomagnetic field sensors 43
   2.2 Magnetic resonance imaging 44
   2.3 Transcranial Magnetic Stimulation 45
   2.4 Instrumentation for monitoring levels of environment exposure to magnetic fields 45
3. NMI measurement capabilities for magnetic quantities compared to health technologies demands 46

Discussion and Conclusion 48

References 49

Author Profile 50

Infusion Pumps Calibration Methods 52

Elsa Batista || Maria do Céu Ferreira

1.Introduction 53
2. Calibration, Error and Traceability 53
3. International and national projects regarding infusion pumps 54
4. Calibration of infusion pumps 55
4.1. Gravimetric method 55
4.2. Comparison method 56
5. Conclusions 57
References 58
Authors Profiles 58

Measuring counted fractions in healthcare 60

L R Pendrill || J Melin

1. Counted fractions in healthcare 61
2. Quality assurance of counted fractions 61
2.1 Quality-assurance standards 61
2.2 Counted fractions 62
3. Restitution of counted fractions and metrological references 63
4. Case studies
4.1 Performance metrics in healthcare services 65
4.2 Cognitive studies of Alzheimer's disease patients 66
5. Metrological item banks
Acknowledgments 67
References 68
Authors Profiles 69

Metrology in Medical Field 71

Baki Karaböce

1. Introduction 72
2. Current Situation in Medical Metrology 74
2.2. International Organizations 77
3. Measurement Parameters and Traceability in Medical Device 80
4. Conclusions 81
References 82
Authors Profiles 82

Statistical methods for a comparative study on Health Metrology 85

Maria do Céu Ferreira

1. Introduction 86
2. Method 87
3. Results 89
4. Discussion 93
5. Conclusions 94
References 95
Authors Profiles 96

Editorial

ANTÓNIO RAMOS PIRES
antonio.pires@estsetubal.ips.pt

MARGARIDA SARAIVA
msaraiva@uevora.pt

ÁLVARO ROSA
alvaro.rosa@iscte.p

Maria do Céu Ferreira
mcferreira@ipq.pt (Editora Convidada)

This special issue of TMQ Magazine is dedicated to Health Metrology. With the main objective of disseminating and sensitizing readers to the topic of health measurements, it was with great pleasure an honor that I received the invitation to take the role of guest editor of such a traditional journal.

The existence of measurements goes back to the origin of human civilization itself, present in a transversal and multidisciplinary way in all areas of society. Measuring instruments, which are part of a multivariate range of activities, allow measurements to be made on a daily basis, at any place and at any time, in an uninterrupted manner.
The technical-scientific context that serves as an insertion into these activities is called Metrology, which is the science of measurement and its applications.

In the health sector, for inherent reasons, measurements and measuring instruments play a key role. Many decisions are based on measurement results, allowing clinical decision to be supported by tangible evidence. The credibility of these decisions, and consequently the difference between a good and a bad decision may depend on whether the information received is correct or not. The metrological approach thus ensures that the results obtained are traceable to metrological standards, contributing to a rigorous and comparable analysis with internationally recognized reference values. This is a fundamental condition for reliable measurements.

To ensure the quality of the measurements, it is necessary to evaluate the performance of process and the measurement system, taking into account all stakeholders. It is known that continuous quality improvement is possible only through the integrated knowledge of all the processes that interfere with the measurement system and its results. There are several factors that may interfere with the measurement result and depend on the characteristics of the instruments, the methodology / procedure used the characteristics of the test sample, the environmental conditions of operation, the handling, etc. Mostly, the factors influencing the measurement result are controllable factors, and a good knowledge of them is therefore imperative.

This issue provides important work on the application of metrology in health, from the presentation of conceptual models for the implementation and development of metrological infrastructures for health services, to data triangulation approaches in the sociological dimension of health measurements, including tools for optimizing available resources. Other articles also deal with methods for calibration of measuring instruments with high relevance in clinical practice, presenting results that promote the improvement of health care quality.

I am sure that the set of articles here presented will aid the reader in understanding and raise awareness to the importance and key value of metrology in health, while also making clear the multiple consequences of f good metrological practices.

May this editorial initiative contribute to a greater contact between interested parties / researchers and professionals of related professional areas, as well as to reinforce the scientific support of metrology as a fundamental pillar of Quality.
Accordingly, my wishful note of gratitude goes to all authors that made this number possible.

O Editor Coordenador
António Ramos Pires

Invited Editorial
Maria do Céu L. S. Ferreira

PAPERS

Patient-centered outcome measurement in psychiatry: How metrology can optimize health services and outcomes

Skye Pamela Barbic
skye.barbic@ubc.ca
University of British Columbia

Stefan J Cano
stefan.cano@modusoutcomes.com
Modus Outcomes

Karen Tee
ktee@foundrybc.ca
Foundry

Steve Mathias
smathias@foundrybc.ca
Foundry

Abstract:

In psychiatry, there is a call for clinicians to use patient centred outcome data routinely at the point of care to help tailor treatment plans to meet the patient's preference and needs. Given that many decisions in psychiatry are constructed from patients' narratives, it is critical that the conceptual, empirical, and measurement structure underlying patient reported outcome measures is robust and patient-centred. Here, we argue for the systematic accruement of patient-centred data in psychiatry to meaningfully enhance the treatment of mental disorders. Specifically, we suggest three crucial considerations for system transformation: (1) the engagement of international patient research partners to conceptualize and prioritize outcomes; (2) the application of modern test theory to develop and evaluate patient-centred outcome measures: and (3) funding allocation accountable to evidence-based services prioritized by patients.

Keywords: Measurement, Metrology, Psychiatry, Rasch

1. Introduction

The treatment of mental disorders is a global health challenge, costing one trillion dollars each year (Chisholm et al, 2016). One in four people will experience a mental disorder at some point in their lives and 450 million people currently suffer from mental health challenges (Peter at al, 2016). As a result, mental illness is the leading cause of ill-health and disability worldwide (WHO, 2013).

Despite established data supporting the impact and prevalence of mental disorders, there remains a colossal treatment gap globally (Peter at al, 2016). It is estimated that 75 % of individuals never receive care, let alone the appropriate care (Meffert et al, 2016). Common metrics of system and personal impact include emergency department utilization, suicide rates, hospital recidivism, and laboratory markers (Walket et al, 2015). Increasingly, it has been suggested that these outcomes alone may not capture the full potential and evolution of psychiatric services (Barbic et al, 2016).

As more people with mental disorders receive care in the community, reintegrating people into productive lives in society is a widespread overarching aim of psychiatric services. Here, we argue for the systematic accruement of patient-centered data in psychiatry to meaningfully enhance the treatment of mental disorders. Specifically, we suggest three crucial considerations for system transformation: (i) the engagement of international patient and family research partners to conceptualize and prioritize outcomes; (ii) the application of modern test theory to develop and evaluate patient centered outcome measures; and (iii) funding allocation that is accountable to evidence-based services prioritized by patients.

2. Engaging patient partners in outcome conceptualization and prioritization in mental health

Patient-oriented research and clinical practice focuses on patient-identified priorities and aligning clinical care accordingly (CIHR, 2016). This “movement” was develop to enhance researcher and clinician understanding of the effects of disease and treatment on patients’ daily lives. Rheumatology and oncology are examples of health fields that have emphasized patient-partners in research in the last decade to improve endpoint outcome measurement (Kirwan et al, 2017). Through their iterative, inclusive consensus processes, outcomes such as participation, function, and fatigue are endorsed targets for treatment and clinical trials in both fields. Systematic approaches to defining patient priorities in psychiatry have yet to be developed. However, psychiatry is not alone. In an editorial published in the British Medical Journal, Angela Coulter (2017) noted that the behavior of measuring what matters to patients is “surprisingly rare”, with a mere 11 % of patient reported outcome measures actually asking patients which outcomes are worth measuring (Coulter, 2017).

In response to this trend, national initiatives, such as Canada’s Strategy for Patient Oriented Research (SPOR) and the Patient Centered Outcome Research Institute (PCORI) in the United States, have been developed to fund research with the intention to help patients feel empowered to make informed decisions about their healthcare choices. As demand increases for accountability of mental health services to be person and family-centered, engagement of patients at all levels of research design are necessary. However, patient engagement in isolation may not be the solution. Unfortunately, many of the outcomes in psychiatry are not directly observable and present systematic challenges for capturing them and characterizing treatment effectiveness. If treatment selection, clinical trials, and/or government policies are to emphasize patient-centered outcomes, a robust approach to measuring them is also needed.

3. The application of modern measurement methods to develop, test, and use patient centered outcome measures in clinical practice

Clinicians frequently use single items (i.e., the Global Assessment of Functioning - GAF) to assess concepts of interest (e.g., function, mood, quality of life) as they appear easy to use. However, single item 'scales' have been repeatedly been shown to be a poor substitute for comprehensive psychological and functional testing (Ramirez et al, 2008). One reason for this is the assumption that a single item 'scale' behaves like a perfect interval ruler, ambitiously claiming to represent a wide concept with one item. Single item 'scales' are actually scientifically limited because they have poor reliability, validity, and responsiveness (Hobart et al, 2007). The limited robustness of this type of scale is also vulnerable to the variability of patient and clinician interpretation. Thus, unlike a ruler that measures length, a single item 'scale' used in psychiatry may cover many different frames of reference that a patient may have. Culture, language, mood, age, gender can all challenge the measurement properties of single item 'scales' and can make comparisons between people, centers, or over time, difficult.

The limitations of single item 'scales' have led to the more frequent use of multiple item scales or questionnaires, in which ratings assigned to each item are combined to give a total score. Combining multiple items to provide a more comprehensive health profile of a patient may seem clinically intuitive. Statistically, this may also be sensible, as evidence supports that combining multiple items reduces random error and enhances scale reliability (Hobart et al, 2007). However, in the case of measurement, statistical significance does not necessarily translate into clinical meaning (and vice versa). Without a robust measurement model underpinning the scoring of multiple item scales, a total score becomes difficult to interpret. This is further complicated because there are many clinical tools that purport to measure the same construct. This makes operationalizing clinical practice and quality improvement impractical. As multiple-item Patient-Centered Outcome (PCO) instruments become increasingly in demand in psychiatry, clinicians and patients are encouraged to learn from advances in measurement to interpret these scales to improve disease management and health outcomes for people with mental disorders (Barbic et al, 2018&2016).

Another area of consideration for psychiatry is the amount of faith invested in the "reliability and validity" of a PCO instrument. Just as significant advances have been made in psychiatric practices (Kidd et al, 2014), notable advances have also been made in the field of measurement. In order to advance the health outcomes of people receiving psychiatric services, it is critical that psychiatry optimizes how measurement advances can improve care and the patient experience (Barbic et al, 2018). New methods in measurement not only provide an opportunity for increased precision in medicine, but an opportunity to use measurement to build a language for service design and provision (Cano et al, 2018). For example, what does a score of 15 (on a range of 0-60) mean in an instrument purporting to measure depression? How does that same score of 15 compare to the total scores generated by dozens of other existing instruments purporting to measure depression?

Interpretation of PCO data has been a longstanding challenge in psychiatry. Communicating results to patients, families, and other health professionals can often be difficult and abstract. This limits the potential to systematically communicate information about how to optimize the recovery and health of people with mental disorders (Barbic et al, 2018). As noted in a recent New England Journal of Medicine Blog, "One of the biggest challenges today is how to design a system that meaningfully engages patients and clinicians for better health" (NEJM, 2017). We argue that engagement cannot occur if key stakeholders do not speak the same language or share the same targets for care.

Fortunately, psychometric approaches, such as Rasch Measurement Theory (RMT) (Rasch, 1980), exist as potential solutions to this problem. RMT has unique properties that are invaluable to fields where outcomes are mostly latent and inferred from the patient (Andrix 2005&1988). These include references for traceability and means of evaluating measurement uncertainty for both patients and items (Pendrill 2014). As well, RMT methods embrace the opportunity for key stakeholder to work together to hypothesize the composition of concepts of interest as linear measures.

RMT methods allow researchers to study the extent to which items form adequately conformable sets to represent clinical concept of interest and map out a clinical hierarchy, to the extent to which specific items can be located to characterize specific areas in a measurement continuum. These then supports hypothesis testing and construct theory development of the clinical concept of interest under study.

RMT methods also allow researchers to study whether the distribution of item locations (or estimates) are independent of the sample distribution of respondents. This is critical for creating measurement tools that are robust for clinical centers, populations, and response shift (Pendrill 2014). The implications for this are profound for all areas of health, notably psychiatry—where most outcomes are subjective, and patients are required to report progress for a long-term treatment period. The application to metrological standards to psychiatry has the potential to foster patient-centered and evidence informed health, to ensure greater quality, accountability, and accessibility of meaningful care (Cano et al, 2018).

4. Patient-centered outcome service delivery and funding allocation

The routine use of PCOs in psychiatry provides an opportunity to help drive how health care is organized, delivered, and funded (Ahmed et al, 2012; Kirwan et al., 2014). As noted above, metrology provides a framework for health sciences to develop a common language between patients, clinicians, researchers and key stakeholders. This form of communication in psychiatry can be used to compare performance metrics of clinical services to evidence-based guidelines or standards. Through benchmarking, there is the potential to learn how well psychiatric interventions perform and align to outcomes that are meaningful to the patient.

Benchmark assessments have been used rigorously by other areas of social science such as Education to monitor progress of stakeholders. Under the assumption that outcomes are derived from rigorous patient-centred conceptual and measurement models, benchmarking can inform the allocation of funding towards the areas of greatest need. This ensures decision makers and policy makers are equipped with the evidence needed to develop plans on how to make improvements or adapt best practices to optimize the health and wellbeing of patients. In this paper, we provide a clinical example of how one Canadian province is applying modern test theory for benchmarking to reconceptualise and fund how services for youth and young adults with mental illness are delivered and evaluated.

5. Example of designing mental health systems with the principles of metrology: Foundry

In 2015, an initiative was approved in Western Canada to develop a new model of care for youth and young adults with mental illness. The purpose of the initiative was to provide "one-stop" services for young people to help them receive the timely and youth-centred support that they need to thrive (Mathias S, 2018). Foundry currently has 120 partnerships across the province of BC and eleven centres that are open or scheduled to open soon. Foundry represents community agencies, government, donors, youth and young adults, and families coming together to improve the wellness of young people in British Columbia. All Foundry centres are branded and present a consistent feel, regardless of geographical location. An example of the two centres and branding is shown in Figures 1a and 1b.

One key ingredient to the success of these centres is consistency of services. It is expected that any young person can walk into any Foundry and expect to find a core set of services that can be offered and a reliable approach to assessment, treatment, and follow-up. To maximize this consistency, our team identified that a youth-centred systematic method of collecting youth-centred data was needed. In response, "Toolbox" was introduced to Foundry.

Figures 1a and 1b. Examples of waiting rooms of two Foundry centers (Kelowna Left, North Vancouver Right), where health data is collected.

5.1 Toolbox

Toolbox© is a data collection platform used at all Foundry centres. The purpose of Toolbox© is to collect youth-centred data to inform service delivery and quality improvement. Toolbox© is an electronic data capture and management system that can capture data from any device (tablet, cell phone, computer, etc...). Toolbox© allows for rapid collection of data to study the extent to which the patient-reported outcome measures are fit for purpose to measure the needs of youth. Toolbox© allows for control over data collection and query management for patient-reported outcomes. Foundry has a dedicated team focused on Toolbox© and its application to each Foundry centre. The team is responsible for ensuring that the data system collects information that is clinically relevant, youth-centred, and robust to measure the needs and priorities of young people across British Columbia. One area of focus for the team is ensuring real-time testing of patient-reported outcome measures (PROMs).

5.1.1. Hypothesis-driven, patient-centered outcome measurement

Well before a PCO instrument is used in Toolbox©, extensive expert (i.e., clinician, youth, family, policy-maker, and psychometrician) feedback is obtained about the candidate measure. First, relevant concepts of interest are identified. Next, a literature review is conducted to identify existing measures to capture these concepts. Once a bank of measures is identified, the content is reviewed by the expert team. The measures endorsed by the team are then subjected to pilot testing at one Foundry centre.

For example, in 2017, Kessler Psychological Distress Scale (K10) (Kessler et al, 2002) was identified as a candidate measure for Foundry to capture youth distress. The desired purpose of the K10 for the team was to evaluate change in distress for all youth accessing Foundry centres. To inform this decision, the K10 was given to 350 young people receiving Foundry services and tested for its psychometric properties. Before we initiated testing, we asked the expert group to hypothesize the ordering of the items (from lowest distress to highest distress) and suggest any anomalies in the data they would expect. Methods guided by RMT were used to test these hypotheses and test the scale properties. Specifically, we tested the ordering of response option thresholds, fit, spread of the item locations, residual correlations, person separation index (PSI), and stability across time.

Tables 1 and 2 and Figure 2 show a summary of the psychometric properties of the K10 for the test population.

Table 1: Traditional psychometric methods-data quality, scaling assumptions, targeting, reliability (n=350).

Psychometric property	Total
Data Quality	
Missing data (%)	0.0
Computable scale scores	350
Scale assumptions	
Item mean scores: mean (range)	2.94 (2.59-3.20)
Item SD: range	1.16-1.36
Targeting	
Mean score (SD)	25.62 (9.67)
Possible score range*	10-50
Observed score range	10-45
Floor/ceiling effect**	<1/<1
Reliability	
Cronbach's alpha	0.90
Mean inter-item correlation	0.01-0.45

* Higher scores indicate greater distress

** Floor effect= % scoring 10 (lowest distress); ceiling effect= % scoring 50 (highest distress).

Table 2: Measures of fit and location (SE) of K10 items. Items are located in order of low (item 1) to higher distress (item 3).

Item	Descriptor	Location [b]	SE	Fit [c]	ChiSq [a d]	Prob
1	feel tired out	-0.505	0.078	3.540	7.98	0.157
8	everything was an effort	-0.439	0.074	-0.397	5.77	0.329
7	feel depressed	-0.357	0.076	-0.735	4.68	0.456
2	feel nervous	-0.355	0.081	0.297	7.43	0.191
5	feel restless or fidgety	-0.228	0.076	1.746	11.34	0.045
4	feel hopeless	0.039	0.077	-2.085	5.30	0.380
10	feel worthless	0.141	0.071	-0.702	4.44	0.488
6	could not sit still	0.491	0.077	1.072	8.82	0.116
9	nothing could cheer up	0.527	0.078	-0.846	13.79	0.017
3	nothing could calm you	0.687	0.079	-0.770	3.71	0.592

[a] Degrees of Freedom (5,248)

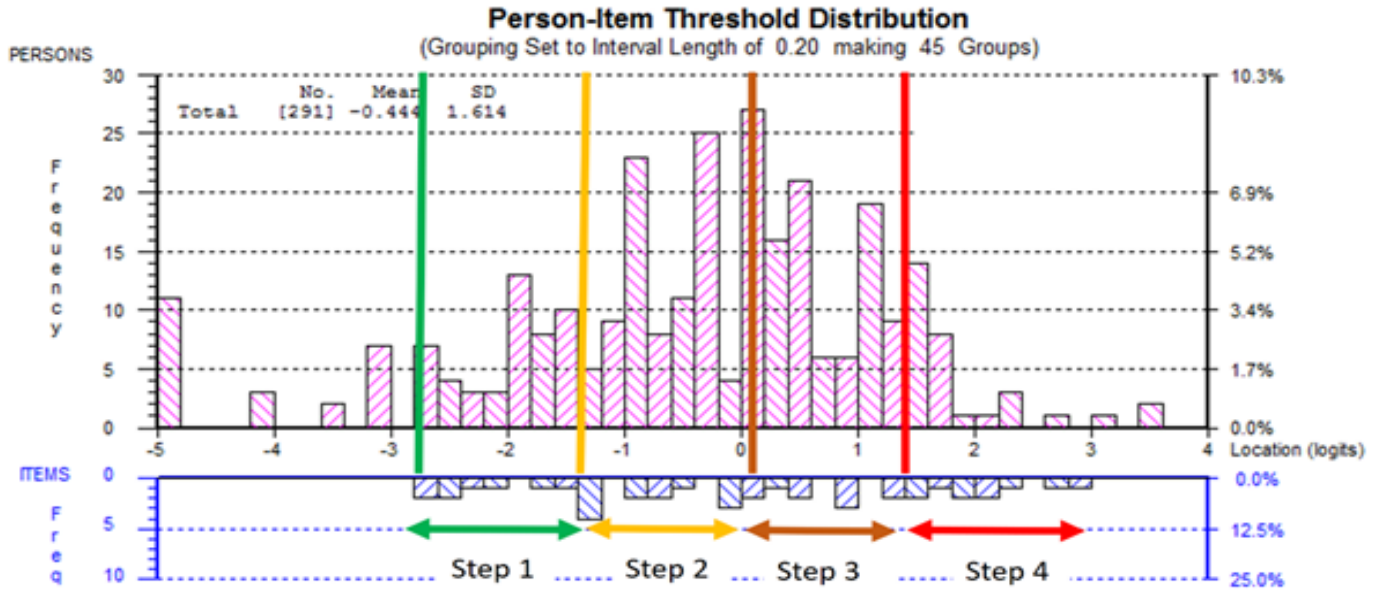
[b] Mean location score obtained for items. In RUMM 2030, the scale is centred on zero logits, representing the item of average difficulty for the scale.

[c] Residual statistics are the standardized sum of all differences between observed and expected values scored over all persons. An item with a residual statistic less or greater than 2.5 warrants further investigation.

[d] Chi-square statistic compares the difference of observed values with expected values across groups with different ability levels (from less to high distress) across the latent trait being studied (distress).

Table 1 specifically outlines the traditional psychometric properties of the tool. Using this method, our team examines the quality of the data, scaling assumptions, targeting (do the items target the population), and reliability (as measure by Cronbach's alpha and inter-item correlations). The information is used to understand the performance of the individual items and how they work collectively to cover the full range of the possible raw scoring options.

Figure 2: Person-item threshold distribution. The distribution of K10 scores for youth are shown by the pink bars- with a range from low (left side) to high distress (right side). An ideal theoretical distribution is from -3 to +3 logits, with a mean of 0 and SD of 1.0. Here we have a mean of -0.44 and SD of 1.6. The blue bars represent the distribution of the item thresholds (-2.7 to +3.0 logits). In an ideal measure the item distribution targets the person distribution. In this example, the targeting of the items to the people is excellent. The cut-off scores of the K10 have been used to partially inform the level of care a person receives based on the cut-off score (Step 1-Step 4). The cut-off scores can be tested regularly to ensure that the level of services is appropriate for the level of distress reported by the patient.



However, we require scales that has interval properties and can be added up to make a total score and one limitation of using traditional psychometric methods is that they are unable to provide the information we need to evaluate this property. Therefore, we used RMT methods to inform the extent to which the items are invariant [a property critical for comparing (1a) a patient over time, (b) patients over time, and (3) different Foundry centres] and were operationalize in the testing context as described by the Expert group (were ordered as expected). As outlined by Rasch himself, "the comparison between two stimuli should be independent of which particular individuals were instrumental for the comparison, and it should be independent of which other stimuli within the considered class were or might also have been compared"(Rasch, 1961: 332). For distress, and other constructs under investigation at Foundry, it is critical that all tools have psychometric properties that are not sample dependent. With the expectation that the number of Foundry centres will grow in number over time, we anticipate that this approach will be critical for regular program evaluation and future PROM development and testing.

5.1.2. Clinically meaningful application of patient-centered outcome measurement

Another area of testing is the clinical utility of the PCO instrument itself. Specifically, what is the extent to which the total score can inform clinical care? As with the physical sciences, our team feels that a total score should produce a meaningful value that is understood by the patient, the clinician, family members and other stakeholders. For example, using methods guided by RMT to understand the underlying measurement model of the K10, we can understand the item hierarchy of the K10 and how items line up consecutively to map a story of the construct. As shown in Table 2 and Figure 2, transferring the location score into a raw score can allow a patient to understand how to interpret the total score. The total score can also become a tangible decision-making tool to map clinical services.

Figures 2 depicts an example for how cut off scores from the K10 are used at Foundry to inform the intensity of evidence-based services. This provides a clear communication tool for patients and stakeholders. A patient can understand, based on their total score, what level of service to expect. This is independent of the centre. This is particularly helpful as well for communicating treatment approaches to other stakeholders involved in the youth's health treatment plan including family members and teachers. The care pathways ensure that a full menu of services is clear to patients and can constantly be evaluated in real time. Foundry is in the early phases of service delivery. However, we anticipate that this approach to program evaluation will allow for benchmarking and the allocation of funding and resources towards innovation and evidence-based services that show clinically meaningful change in a patient.

The rigorous approach to measurement allows ongoing opportunity to identify and understand practices, methods, and processes at Foundry. As an organization, this allows the potential to customize practices to individual youth and each Foundry centre. We also anticipate that benchmarking, based on meaningful measurement, will allow for more efficient use of resources through the identification and implementation of best practices modified to local Foundry centres and communities. Benchmarking will allow our team to focus on comparing internal processes and innovation and compare those from other initiatives in Canada and beyond. As such, data are used in a clinically meaningful way as a communication mechanism for continued improvement and learning, as well as networking with like-minded organizations to optimize the health and well-being of young adults with psychiatry disorders.

6. Conclusion

Patient-centred outcome measurement in psychiatry is essential for advancing outcomes and delivering high quality services to patients and their families. The capacity to utilize metrological standards in psychiatry should be a fundamental aspect of health service design, delivery, and evaluation in psychiatry and mental health. To do so, partnerships between clinical leaders, patients, and measurement experts are essential. Patients receiving psychiatric services have the right to the highest quality of care and to understand how their health outcomes translate to evidence-based services. Entry points to action include a formal linkage with the metrology and mental health communities. Strengthening the measurement-health linkage will improve international quality mental health services, interventions, and patient experiences.

References

Chisholm, D et al., (2016), Scaling-up treatment of depression and anxiety: a global return on investment analysis. *Lancet Psychiatry* **3**, 415-24.

WHO-World Health Organization, (2013), *Mental Health Action Plan 2013-2030* .Geneva, Switzerland.

Patel, V et al., (2016), Addressing the burden of mental, neurological, and substance use disorders: key messages from Disease Control Priorities. *Lancet* **387**, 1672-1685.

Meffert, SM et al., (2016), Novel implementation research designs for scaling up global mental health care: overcoming translational challenges to address the world's leading cause of disability .*Int J Ment Health Syst* 10 19.

Walker, ER et al., (2015), Mortality in mental disorders and global disease burden implications: a systematic review and meta-analysis. *JAMA Psych* **72**, 334-341.

Barbic, SP et al., (2016), The application of Rasch measurement theory to psychiatric clinical outcomes research: Commentary on Screening for depression in primary care. *BJPsych Bull* **40**, 243-244.

CIHR-Canadian Institutes for Health Research, (2016). *Strategy for Patient-Oriented Research.* Canada.

Kirwan, J R et al., (2017). Emerging Guidelines for Patient Engagement in Research.

Value in Health **20**, 481-486.

Coulter, A (2017), Measuring what matters to patients. *BMJ* 356, j816.

Ramirez, A et al., (2008). Global Assessment of Functioning scale (GAF), further evaluation of the self-report version. *Eur Psych* **23**, 575-579.

Hobart, J et al., (2007), Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. *Lancet Neuro* **6**, 1094-1105.

Barbic, SP et al., (2018), Patient-centred outcome metrology for healthcare decision-making. *J. Phys. Conf. Series* **1044**, 1-6.

Kidd, SA et al., (2014), Mental Health Reform at a Systems Level: Widening the Lens on Recovery-Oriented Care. Can *J Psych* **59**, 243-249.

Cano, SJ et al., (2018), Patient-centred outcome metrology for healthcare decision making.

J. Phys. Conf. Series **1044**.

Catalyst, NEJM (2017), Hardwiring Patient Engagement to Deliver Better Health (https://join.catalyst.nejm.org/events/archived-hardwiring-pe-2017/register/on-demand).

Rasch, G (1980), *Probabilistic models for some intelligence and attainment tests.* Copenhagen, Demark.

Andrich, D (1988), *Rasch models for measurement.* Beverley Hills, United States.

Andrich, D (2005), *Stat Behav Sci* **4** 1698-1707.

Pendrill, L (2014), Man as a measurement instrument.t *J Meas Sci* **9**, 24-35.

Ahmed, S et al., (2012). The Use of Patient-reported Outcomes (PRO) Within Comparative Effectiveness Research: Implications for Clinical Practice and Health Care Policy. *Med Care* **50**, 1060-1070.

Kirwan, J et al., (2014), Updating the OMERACT filter: implications for patient-reported outcomes.

J Rheumatol **41**, 1011-5.

Mathias, S (2018), Foundry: Where wellness matters. URL: www.foundrybc.ca.

Kessler, RC et al., (2002), Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychol Med* **32**, 959-976.

Rasch, G (1961), *On general laws and the meaning of measurement in psychology.* Berkeley, USA.

Authors Profiles:

Skye Barbic has received a Ph.D. from McGill University- Montreal, Canada in 2013. She is now an Assistant Professor in the Department of Occupational Therapy at the University of British Columbia. – Vancouver, Canada and the Research Lead for Foundry. Her research interests are in the areas of Health Measurement, Youth Mental Health, and Rasch Measurement Theory.

Stefan J. Cano is a Chartered Psychologist and Associate Fellow of the British Psychological Society. He 24 years' experience in patient centered outcome measure development and psychometric research. Having previously worked extensively in qualitative and quantitative research projects, Stefan's main area of interest is in developing, applying, and improving mixed methods psychometric research in clinical studies and therapeutic trials.

Karen Tee received a Ph.D. from Simon Fraser University, Burnaby, Canada. She is a Clinical Psychologist and the Director of Service Innovation for Foundry in Vancouver, Canada, and has experience in youth mental health both in clinical operations and direct service delivery for the past 25 years.

Steve Mathias received an MD from the University of British Columbia – Vancouver, Canada. He is the Executive Director for Foundry (foundrybc.ca) in Vancouver, Canada and the Head of the Department of Psychiatry at St. Paul's Hospital. His research interests are Integrated Health, Youth Mental Health, and At-Risk Youth.

A Nondualist Social Ethic: Fusing Object and Subject Horizons in Measurement

William P. Fisher, Jr.
wfisher@berkeley.edu
University of California, Berkeley

Abstract:

Is it possible for measurement to satisfy both the need for global unity and shared human values, on the one hand, and the need to respect and value individual human uniqueness, on the other? Can humanity figure out new ways of enacting self-understanding that recount its past, present, and future in terms that not only better enable everyone to feel part of something larger than themselves but which also liberate them in free creative expressions of joy for life? Any viable social ethic has to begin from the paradox of attempting to integrate the opposed poles of human totality and human singularity. Even when one conceives of a social ethic in terms of a system, the end result demands respect for a complex combination of harmony and dissonance that is better understood in terms of systems of discontinuous systems, or meta-systems. Taking language as a model of integrated subject-object unities sets up possibilities for coherent multilevel measurement information infrastructures. Health care, education, management, and other areas in which ordinal counts, percentages, and ratings typically are treated as measures stand to benefit the most from realizations of this kind of nondualist social ethic and its realization of meaningful quantification.

Keywords: complexity; information infrastructures; measurement; nondualist philosophy

1. Dualism and Nondualism

Various kinds of dualism are described or assumed in the history of philosophy, such as mind and body, idealism vs materialism, feelings vs facts, observer vs observed, determinism vs randomness, or even people vs things (Pickering, 2010). Searle (2010) contends that dualist accounts of consciousness primarily involve the opposition of idealism and materialism; he further holds that idealism is dead, and that the only remaining philosophical approaches to accounting for consciousness are dualism and materialism. In contrast with this perspective, others, from Whitehead (1911, p. 51) and Freud (1920) to Hayek (1948, p. 54) to Gadamer (1989) to Rowlands (2010), point out that subjective unconscious and subconscious factors unavoidably affect experience of the objective world of things. A key issue concerns the way objective language and linguistic technologies like alphabets, phonemes, grammatical rules, text, and readable instruments represent things in the world in ways that embody shared externalized cognitive processes. These technologies effectively prethink the objective world as something communicable and shared. In so doing, however, linguistic technologies also embody subjective factors of culture and history. This contrast of objective reproducible factuality and subjective subconscious intuition is the context in which the present work approaches a nondualist social ethic.

Even as far back as 1895 (Tarde, 2012, p. 5), the abyss between objective and subjective forms of understanding that had long been judged unfathomable was characterized by the "bravest" of Leibniz's philosophical heirs as mere appearance or was denied to exist at all. This denial of dualism, Tarde says, was substantiated by the various sciences' successes in producing "a prodigious multiplication" of unified subject-object monads throughout the economy.

Little attention, however, has been accorded over the course of the intervening 120 years and more to contrasts of the pragmatic implications of dualist and nondualist philosophies for methodology in psychology and the social sciences. Modern Western science typically understands itself as dualistic in the sense of assuming the existence of an independent objective world separate from the idiosyncratic subjective experience of an observer. Method in this context is taken to be a matter of following rules in a prescribed fashion to obtain a predetermined end product.

Various strands of criticism of this subject-object dualism arrive at the realization that understanding inevitably requires a mutual implication of subject and object involving their joint participation in a common flow of experience (Dewey, 1925/1981, 2012; Gadamer, 1989, pp. 460-463; Galison, 2008; Haraway, 1996; Harding, 2008; Heidegger, 1962, 1967; Kuhn, 1970; Nersessian, 2012, 2015; Nussbaum, 2013; Overton, 2015; Toulmin, 1953; Whitehead, 1925). In Gadamer's (1989) terms, insofar as subjectivity is disciplined by strictly following objective methodical rules, truth is inaccessible, and insofar as new true facts are recognized, method is abandoned, leading Ricoeur (1981, p. 60) to suggest that Gadamer's book, Truth and Method, might better have been entitled, Truth OR Method. Kuhn (1970) similarly found that scientific textbooks typically rewrite history, presenting method recast in terms of the desired end product, where a review of the journal articles reporting developments in the relevant area of research reveal a much less straightforward account of events.

Recognition and acceptance of the role metaphysical subjectivities inevitably play in science can emerge in unexpected quarters, as in Campbell's perspective on physics (1920, pp. 10-12, 231-233), but neither this acceptance nor the aptness of the critique has resulted in the widespread adoption of nondualist methods. Instead, the natural and social sciences have both largely continued enacting and tolerating a variety of modern ontological and epistemological divides between what scientists say they do, and what they actually do (Harding, 2008; Latour, 1987, 1993, 2013). Postmodernism functions as a kind of disappointed form of modernism (Latour, 1991, p. 17) because it does not follow the methodological implications of its critiques through to amodern (Latour, 1990, 1993) or unmodern (Dewey, 2012) enactments of conceptualizations better able to extend the playful model-based reasoning practices of everyday thinking (Nersessian, 1996, 2002, 2006, 2008, 2015) in new directions.

Differences between theory and practice in scientific method can be characterized in large degree as differences between dualistic and nondualistic accounts. Human behavior and cognition are increasingly conceptualized in nondualistic terms as a function of socially embedded and distributed knowledge technologies (Dewey, 2012; Hutchins, 2010, 2014; Latour, 1987; Nersessian, 2006, 2012, 2015; Sutton et al., 2010). These technologies include language's alphabets and numbers, scripts, grammars, vocabularies, etc., and the media in which language is embodied and distributed, such as phonemes, books, digital fonts, etc. As was recognized to various extents by Whitehead (1911, p. 61), Hayek (1948, p. 88), Tarde (1903a, 1903b; Latour, 2002, 2010; Latour & Lépinay, 2009), and Platt (1966), what Mach (1919, p. 481) called the labor-saving effects of the economy of language embody and distribute functional sign-thing-concept assemblages when new things come into words via a collective process not directed by any individual or group. Much of the puzzle as to how systematic methodologies might be formulated and deployed in practice follows from the leaderless distributed processes enacted in concept formation (Hutchins, 2010, 2014; Nersessian, 2008, 2015). Indeed, Hayek (1948, p. 54) considered this puzzle to be the central question of all social science.

History in general (Braudel, 1992), and the history of science in particular (Cohen, 1985; Holton, 1988), suggest that fundamental conceptual and practical changes almost never occur because of persuasive arguments showing one idea or way of doing something is more efficient, aesthetic, meaningful, or useful than another. Kuhn (1970, p. 151), for instance, recounts Planck's recognition that "'a new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it."

Closer consideration leads to recognition of an important role for technology in conceptual change. New theories and data emerge in the wake of new technologies affording shared and widely available capacities for reproducing controlled effects (Bud & Cozzens, 1992; Hankins & Silverman, 1999; Latour, 1983, 2005; Price, 1986; Schubert, Sydow & Windeler, 2013). Artists and artisans played significant but often unrecognized roles in engineering these technologies (Smith, 2006). Similarly, the roles of technicians in facilitating the production of the objects of study in science were systematically ignored by historians until recently (Schaffer, 1992, p. 23; Shapin, 1989). But it is nonetheless the case that "thermodynamics owes far more to the steam engine than ever the steam engine owed to thermodynamics" (Price, 1986, p. 240). Repeated absorption of researchers' attention into the reproduction of controlled effects leads to the formation of ideas about that experience and the negotiation of a new language. The medium is the message in McLuhan's famous expression, "the symbol gives rise to thought" (Ricoeur, 1967), technology has historical and ontological priority over science (Ihde, 1983, 1991), the victory of method over science characterizes the nineteenth century more than the victory of science itself (Nietzsche, 1967, Nietzscke, 1967), and "men have become the tools of their tools" (Thoreau, 1854, p. 51). Kuhn (1977) accordingly notes that:

> Of the nine pioneers who succeeded, partially or completely, in quantifying conversion processes, all but Mayer and Helmholtz were either trained as engineers or were working directly on engines when they made their contributions to energy conservation. Of the six who computed independent values of the conversion coefficient, all but Mayer were concerned with engines either in fact or by training. (p. 90).

Human life has always been technologically embedded (Ihde, 1990). Language itself is a knowledge technology embodying concepts that do not have to be worked out by each of us, individually. Using signs to represent things is not just an effect of social life; social life is itself shaped in terms of symbolic functions (Ricoeur, 1981, p. 219; Latour, 2005), which constrain the ways in which each culture's metaphors uniquely shape concepts and the relationships they inform.

Technologies from words to thermometers to automobiles to electronics embody what Gadamer (1989) described as a fusion of subject and object horizons (Butler, 2016; Heelan, 1991). Technologies absorb subject and object together in a playful movement encapsulated in a ready template for easy copying and imitation. This capacity to fuse subject and object horizons in the technological medium is the means by which language performs the advance work of prethinking the world for us, as Gadamer (1989, pp. 429-430) puts it, and so provides an economy of labor-saving utility (Mach, 1919). Words embody a referential function that integrates the representation of objects in the world with subjective

conceptualizations and lived experiences. This capacity for portable encapsulation of a ready-made integration of object and subject horizons is what makes it possible for individuals to rarely have more than rudimentary grasps of how any technology works. Even everyday technical objects like alphabets, light switches, and automobiles involve origins and complexities scarcely imagined by most people using them.

Individual understandings of things may vary but insofar as communication is achieved, it comes about in terms of the ways words enact a given subject-object fusion. We become habituated to the subjective stance evoked by any given technology, such as the meaning of a word in a particular language, and so perform and inhabit that particular cultural space when we give over our subjectivity to an experience of an object mediated by the word in use. This fusion of horizons is the crux of nondualistic thinking and acting, and provides a much-needed basis for methodological advancement in psychology and social sciences.

Contrary to what might be expected, this fusion does not deny the existence of a separate independent world. Instead, it recognizes that "the problematic of objectivity presupposes a prior relation of inclusion which encompasses the allegedly autonomous subject and the allegedly adverse object" (Ricoeur 1981, p. 105). The separation of things from words follows from and is made possible by their prior unity in a shared conceptual frame of reference. The modern dualist denial of the mutual implication of subject and object has covered over and hidden the means by which nature came to be understood mathematically (Husserl, 1970, pp. 23-59). The need to recover an understanding of how science became mathematical necessitated postmodern efforts focused on finding ways to understand "things as they show themselves before the work of abstraction and theorizing has carved out a language of fixed essences for them removed from human praxis, history and culture" (Heelan, 1994, p. 369).

A different perspective on causality is opened up by this distinction between experience caught up in a flow of unnoted and undifferentiated subject-object horizons, on the one hand, and experience informed by familiar fusions of subject-object horizons, on the other. The difficulty or agreeability of questions asked on an instrument of psychological or social measurement has been held to cause an associated correct/incorrect or agreeable/disagreeable response in proportion to the ability or attitude of the respondent (Stenner, Fisher, Stone & Burdick, 2013). An alternative perspective follows from the mutual implication of subject and object, and offers a different kind of structural explanation. Instead of sequential or consecutive cause-effect events that do not share an inner logical connection, fused subject-object horizons imply a simultaneous and correlative belonging together of cause and effect as a unitary phenomenon (Ricoeur, 1981, p. 219). The shift of focus here has extensive implications for social analysis, as Ricoeur notes and as has been the topic of debate at least since the time of Tarde and Durkheim (Candea, 2010; Latour, 2005, pp. 14-16; Latour & Lépinay, 2009; Hayek, 1948; Platt, 1966, pp. 47-49).

Because of the way things come into words via leaderless, collective social processes, and because of the way speakers, listeners, readers, and writers are absorbed into the play of language games, Gadamer (1989, pp. 101-134) asserts that play is the fundamental clue to authentic method. To understand an utterance or a text is to be drawn into an event in which meaning asserts itself; we are captivated by the meaning (Gadamer, 1989, p. 490). This captivation takes place in the same way that the play of sporting, musical, historical, or dramatic performances pulls players and spectators alike into an experience in which one thing follows another of its own accord, and is always already in play by the time it is apprehended.

Focusing on human play as being of the same ontological domain as the play of animals, of waves on the beach, and of light through the leaves offers the potential for resituating humanity again as part of the natural world. Method, at root, is a following along after (meta-) the experience of things themselves along the path (-odos) they take (Heidegger, 1991, p. 63; Gasché, 2014). The structure of play in language games exhibits the economy of language (Mach, 1919, p. 481) in that it lessens the burden of taking the initiative in concept formation (Gadamer, 1989, p. 104) and the determination of concepts expressed in recognizable signs. With no reference to Gadamer's analysis of play as the clue to method, Nersessian (1996, p. 542) similarly points toward child's play as analogous to scientific experimentation in that both are concerned with "understanding and explanation by means of theory development."

These two aspects of nondualism, the embodiment of meanings in words that consistently refer to things in the world across speakers and writers, and the authentic experience of playful captivation in the flow of meaning, provide the focal points of methodological interest (Fisher, 2004; Fisher & Cavanagh, 2016). Given that scientific inferences are so suspended in language that we cannot tell up from down (Bohr, in Petersen, 1968, pp. 187-188), and that scientific thinking and language are extensions of everyday thinking and language (Bohr, 1963, p. 9; Huxley, 1862, p. 57; Einstein, 1954, p. 290; Nersessian, 2006), how might we devise new answers to Hayek's (1948) "central question of all social science"?

> How can the combination of fragments of knowledge existing in different minds bring about results which, if they were to be brought about deliberately, would require a knowledge on the part of the directing mind which no single person can possess? To show that in this sense the spontaneous actions of individuals will, under conditions which we can define, bring about a distribution of resources which can be understood as if it were made according to a single plan, although nobody has planned it, seems to me indeed an answer to the problem which has sometimes been metaphorically described as that of the 'social mind.' (p. 54).

Hayek's (1948, pp. 6-13) sense of the true individual then projects a social whole larger than the sum of its parts, none of which has complete information or is perfectly rational. Hayek does not follow through in the way Tarde does to a sense of individual-level quantification. But both Hayek and Tarde tap deep common sources in their independent approaches to understanding matters that today are conceptualized in terms of individual- and agent-based models of complex adaptive self-organization (Arthur, 2015; Railsback, 2001; Fisher, 2017).

In science, the key issue concerns the embodiment in standardized instruments of real phenomena in the world that do not have to be rediscovered and reinvented, but are made available to everyone trained in the language and connected in the metrological traceability network. Latour's (1987, 1998, 2005) emphases on the ways in which these kinds of networks and ecosystems inform rethinking of the social harmonize with his appreciation for Tarde's monadology (Latour, 2002, 2005, pp. 14-16; Latour & Lépinay, 2009), and are fundamentally nondualistic (Murdoch, 1997).

The larger problem pointed at by metrological networks is, however, how things come into language in the first place by means of the partial, local knowledge possessed by individuals. What unmodern/amodern nondualist alternative is there to the modern/postmodern dualist top-down imposition of centrally controlled, bureaucratically manageable universals and uniformities? How do things in language emerge from the bottom up via particulars and processes not controlled by any individuals or groups? And how do these processes result in continuous, navigable, idealized abstract structures at the same time they support local, concrete, malleable flexibility and innovation (Scott, 1998; Star & Griesemer, 1989; Star & Ruhleder, 1996)? This is the matter of most concern in figuring out what it might mean to take language as a model (Scott, 1998, p. 357) for information infrastructures in health care, education, and other areas.

The problem of nondualistic method in psychology and the social sciences hinges on how the gift of emergent meaning can be embodied in common languages of measurement that integrate locally situated practical knowledge supporting creative improvisations with abstract ideals providing continuity and navigability. The opportunity and challenge global humanity faces concerns the commonly offered, unasked for, and as yet only rarely accepted gift of repeatedly identifiable invariant constructs: learning progressions, healing trajectories, and developmental sequences that have been shown over decades of research and practice to remain stable across assessment and survey data sets, and across clinics, classrooms, time, and space (Bond, 2008; He & Kingsbury, 2016; Dawson, 2004; Fisher, 1997a, 1999; Rasch, 1960; Wilson, 2005; etc.).

The opportunities at hand for advancing theory and practice are analogous to the relation of the steam engine and thermodynamics in the history of physics. Experiencing and re-experiencing, and expressing and re-expressing, the same constructs across samples, instruments, researchers, etc. sets the stage for new things to come into words, words capable of mediating new relationships within and across locally situated forms of practical knowledge. Meaningful representation of these repeated patterns in common languages has to date been rare but not nonexistent (Fisher & Stenner, 2016). Decades of research and practice have firmly established the fact that persistent and repeatable constructs can be identified and measured in common quantitative units defined in terms of simultaneous, conjoint relational structures across samples of persons and questions (Rasch, 1960, 1977; Andrich, 1988, 2010; Bond &

Fox, 2015; Engelhard, 2012; Fisher & Wright, 1994; Luce & Tukey, 1964; Wilson, 2005; Wright, 1968, 1977, 1999; Wright & Stone, 1979, 1999). Methodical play provides media for this act of things themselves experienced in thought, to adopt Gadamer's (1989, pp. 460-464, 474) felicitous phrase, when it embodies constructs in scales giving form to their own self-organized expression (Fisher, 2004, 2017).

If the features of the questions asked in research studies can be well enough understood to enable a theory-based explanatory synthesis of the scale, a thorough conceptual grasp of the construct is demonstrated (Dawson, 2004; Embretson, 2010; Fischer, 1973; Stenner, et al., 2013; Wilson, 2005). The encapsulation of this theory and the lessons learned from data in an instrument scaled in a standard unit then comprises a nondualistic whole ready for deployment in a distributed metrology network capable of serving as the external scaffolding for shared cognition and organizational memories (Sutton, et al., 2010; Hutchins, 2010, 2014).

That is, what Gadamer calls the activity of the thing itself manifests as invariant patterns of progress in learning, sequences in development or degeneration, or trajectories in healing. These are mapped by and embodied in the instrument. Applications employing the instrument take advantage of the economical prior determination of a conceptual domain validated via experiment and theory as sufficiently and invariantly represented in the reading of the tool's standardized terms. The resources invested in developing tools capable of lifting "the burden of taking the initiative" (Gadamer, 1989, p. 104) are paid back many times over in the ways communication is facilitated, and in the way unexpected anomalies provoke surprise and further investigation.

The tool then embodies the solution to the problem being studied or managed. The way forward from the position indicated by the measure is built into the interpretation. In contrast with the modern assumption that problems are inherently independent of their solutions, the nondualist unmodern perspective sees the mutual implication of problems and solutions in forms of life occupying various niches in sociocognitive ecosystems. This emerging perspective and sphere of practice follows from the philosophical importance of recent work relating psychometric calibrations of invariant interval scales from ordinal data to metrological definitions of unit standards, uncertainty evaluations, and traceable instrument conformity assessments (Mari & Wilson, 2014; Maul, et al., 2018; Pendrill, 2014; Pendrill & Fisher, 2015).

2. From individual experience to institutions and systems

Nondualistic philosophy is usually considered a matter of individual spiritual experience, with its primary values being an appreciation of wholeness that informs an enhanced sense of well-being, compassion for others, and evening out the ups and downs of life. Most, if not all, formal religions have traditions or perspectives asserting one or another kind of undifferentiated holistic knowing, though these can vary markedly in how they characterize that knowing, and in the quality of the experience obtained (Yandell, 1994). These variations in the quest for what are considered transcendental experiences of purely spiritual wholeness are then sometimes vulnerable to criticisms (from within these traditions as well as from without) that they are unaware of the role of language in mediating consciousness, conceptualization, and understanding (Mohr, 2000).

The consequences of these critiques include the idea that enlightenment may be less a matter of individual insight and effort than a product of the concepts embodied, distributed, and made available socially. Instead of opposed and self-righteous varieties of religious persuasion, perhaps the intrinsic locus of enlightenment could be systematically relocated at a social level of organization in the form of distributed networks of interconnected linguistic knowledge technologies. This could be possible should the coordination of collective processes of distributed cognition embody nondualistic expressions of human compassion and caring.

Gilligan (1982), Heidegger (1962), Noddings (1984) and Ruddick (1989) focused on care as the defining human quality needed to clear a space for shared understanding. Habermas (1995, p. 199) similarly stressed considerateness for shared vulnerability as fundamental to social life and communication, while Irigaray (1984) found a source of fundamental meaningfulness in lovers' fecund gifts of life. Gadamer's (1991, p. 61) examinations of humanity's captive enthrallment with the allure of meaning and beauty begin from care for the unity and sameness of the object of the conversation. Ricoeur (1974, pp. 95-96) concurs, saying that "philosophy is entirely defined by the desire for meaning, by the choice in favor of coherent discourse." In all of these, we begin from care, desire, and love:

> For, Plato argues, in love we are both ignorant and wise, finite and infinite, possessing and lacking. The lover, in longing for his beloved, cannot be said to possess nor to lack what he desires, since he would not love if he totally lacked, nor would he be able to desire if he totally possessed. It is through this extraordinary phenomenon of love that we thereby come to understand how meaning can be thought about. For in thinking about meaning, we neither fully possess the perfect form of meaning (e.g., the ideal state), nor are we totally unaware of it (Gelven, 1984, p. 132).

Reflecting on the lessons of Diotima in Plato's Symposium, Socrates accordingly said that we are enthralled with meaning in the same way a lover is captivated by the beloved. The physical experience of embodied desire is extended in language in such a way that, as Gadamer notes, in thinking, speaking, listening, reading, and writing, we are as possessed by the desire for understanding as lovers are absorbed in mutual fascination.

The experience of love teaches us how to understand meaning in a way applicable to problems of measurement (Fisher, 2018). Integrating a tolerance for ambiguity, in the form of a suspension between an unattainable abstract ideal and the available concrete evidence, appears in pragmatic terms, for instance, in geometry's representation of line segments of undrawable lengths, like the square root of two, as well as in measurement theory and practice, where it is commonly recognized that models are not meant to be true, but to be useful (Box, 1979, p. 202; Cartwright, 1983; Rasch, 1960, p. 37-38; Rasch, 1973/2011). Social studies of science have similarly arrived at analogous appreciations of the way concrete local forms of knowledge resonate with and diffract higher order abstractions (Berg & Timmermans, 2000; Haraway, 1996; Star & Griesemer, 1989; Star & Ruhleder, 1996). Latour (2011) accordingly recommends caring for our technologies the way we care for our children. These everyday ways of separating and balancing locally situated concrete improvisations and abstract general ideals offer the opportunity for facilitating improved management of outcomes in education, health care, social services, natural resource management, etc. (Fisher, 2018; Fisher, Oon & Benson, 2018; Fisher & Stenner, 2018; Fisher & Wilson, 2015).

A key direction is indicated by Latour's (2005, p. 14; 2010) remarks on Tarde's perspective on the way quantification relies on both aggregate and individual information. This same concern with the emergence of coherent, self-organized patterns of shared meaning from individual responses to test, assessment, and survey questions pervades contemporary psychological and social measurement (Rasch, 1960, 1977; Andrich, 1988, 2010; Bond & Fox, 2015; Engelhard, 2012; Fisher & Wright, 1994; Wilson, 2005; Wright, 1977, 1999; Wright & Stone, 1979, 1999). Applications of Rasch's (1960, 1977) models of measurement are well established as fulfilling Tarde's belief "that we could invent the instrumentation for capturing the inner quantification of individual entities" (Latour, 2010, p. 150).

Indeed, from the nondualist perspective on fused subject-object horizons, "what appears most qualitative is actually where the greatest numbers of calculations are being made among 'desires' and 'beliefs'" (Latour, 2010, p. 154). A large body of existing research thus agrees with Tarde that the matching of the desire for meaning with opportunities for making it, and the matching of belief in one's abilities with appropriate challenges in the world, are "the locus where we should be best able to quantify" (Latour, 2010, p. 154). Quantification based on probabilistic models furthermore agrees with Tarde that there are no individual atoms that aggregate into wholes equal to the sums of their parts. Instead, wholes are always greater than the sums of their parts. This is because they are open systems modeled in stochastic terms as pertaining to infinite populations, not defined totalities (Fisher, 2017). Custom-tailored approaches to measurement are then capable of adaptively adding and dropping questions and answers across conditions without compromising the unit of comparison, which is understood as obtaining within a given range of uncertainty (Barney & Fisher, 2016; Lunz, Bergstrom & Gershon, 1994).

The integration of (a) unique patterns in individual-level concrete responses to questions with (b) an abstract aggregate pattern that persists across samples and sets of questions provides the substantive means by which things can be brought into language as fused subject-object horizons (monads). The paradoxical simultaneity of abstract ideality and concrete flexibility that is characteristic of natural language's capacity for both continuity and innovation has long been incorporated into reports tailored to the needs of different stakeholders (Wright, Mead & Ludlow, 1980; Chien, Linacre & Wang, 2011;

Masters, Lokan & Adams, 1994; Mead, 2009; Wilson, 2005, 2011; Fisher, Oon & Benson, 2018). This capacity to separate the content, complexity, and type of information presented to different stakeholders needs to be systematically built into information infrastructures. Failure to do so results in continued incoherent and unmanageable disconnections between the dictates of abstract idealization and local demands for flexible adaptation (Scott, 1998; Star & Ruhleder, 1996).

The point of contrast between dualist and nondualist methods hinges on caring about the way creative innovations are understood to bring new things into words. In research methods employing tests, assessments, surveys, and questionnaires, the predominant and widely criticized (Michell, 1986, 2000; Maul, 2017) approach to measurement is popularly conceived to be the assignment of numbers to objects according to a rule. The objectivity of these assignments is held to be contingent on the repeatability of the operations involved. This objectivity is deemed sufficient to the task of measurement, defined in terms of descriptive statistical modeling, and not in terms of prescriptive modeling of invariant unit definitions. Subjective influences such as feelings and biases are considered undesirable and efforts are made to omit or remove them from the investigation.

The representations produced in measurement of this kind, however, commonly involve numbers that do not stand for anything that adds up in a way that is meaningful in terms of a repeated interval unit. The assumptions of this modern subject-object dualism are that the medium is not the message, that symbols do not give rise to thought, that play is not the clue to method, and that society is not fundamentally symbolic. Society is instead assumed to compose messages using language instrumentally, as a tool subject to the will of individuals able to use it as they see fit, applying it to objects in the world that are not conceptually entangled with words in complex ways.

But consider what happens when numbers are assigned to objects according to a rule defined by questions on a test or survey. Numeric scores are derived from sums of ratings, or from counts of correct answers. The meanings of these scores are inextricably tied to the particular questions asked. If a question is skipped by someone, the resulting score is not comparable with the scores obtained from people answering all of the questions. In the same way, if one or more questions on the tool are changed, dropped, or added, new scores are not comparable with old ones.

This artificial scheme stands in marked contrast with the way languages evolve over time. Words come into and exit from languages via group-level social processes not directed by anyone (Gadamer, 1989, p. 463; Oudeyer, 2006). Words take their positions in language in relation to all the other existing words, and acquire their meanings via interactions that define how they differ and correspond. The end result is a word that is generally meaningful across individual instances of the objects it stands for, and across the subjects referring to it. The word is furthermore accessible to anyone skilled in the language, and in a way that does not require understanding how the word, its pronunciation, its spelling, or its conceptual determination came about. Scientific research extends this process by establishing reproducible experimental and theoretical relations that can be conveniently packaged in a technology communicating new knowledge. The generality of scientific terms follows from the efforts made to establish firm associations between repeated productions of a phenomenon and representations of it.

It is often said that innovation is a team sport precisely because technological creativity advances civilization by making it possible for people to execute operations they do not understand (Whitehead, 1911, p. 61). Different perspectives on the same thing brought to bear by people with different areas of expertise, combined with a common language for efficient communication, are necessary to the creative dynamics of fields and organizations (Woolley & Fuchs, 2011). In the work of these organizations, human subjectivity is less a source of error that has to be identified and removed from research and decision making than it is an inevitable part of human existence. Instead of continuing to pursue the futile modern project of becoming as fully conscious and free of prejudices and biases as possible, we ought instead focus on going with the flow of subjectivity. In this scenario, feelings and prejudices are put on the table for comparison against objective data, and for possible refutation by others' subjective perspectives. As Wright (1958) put it, objectivity may be the royal

road to reliable knowledge about the external world. But when we are trying to understand ourselves and how we learn, scientific objectivity does not seem to be enough. Perhaps we need to embark on another road, one that is more subjective. (p. 368).

Wright (1958) continues, saying:

> Occasionally the unavoidable impact of subjectivity in research is explicitly recognized. But then the influence is usually acknowledged only as a source of error. Efforts are focused on trying to rid the experiment of its subjective aspects in order to approach the hopefully scientific goal of objectivity. But these efforts at objectivity sanforise right out of the research the very data that, it seems to me, are most likely to help me out of my dilemma. Instead of trying our best to get rid of the subjective aspects of our research, we might better try our best to harness our subjective experience in a way that would allow us to sort out and make the most of its contribution. (p. 369).

To begin harnessing subjective experience, Wright (1958, p. 371) "calls for a special kind of inner act," an empathic identification with another person. This seems to be an instance of the Golden Rule (treating others as we would like to be treated), where the goal is to entertain the perspective an examinee or survey respondent might take when responding to questions on a particular topic. Wright suggests taking on the goal of planning "a course of action that includes a feeling for what moves the child" taking a test, or the person responding to survey questions (Wright, 1958, p. 371). A feeling for what moves a student will likely involve sensitivity to developmental sequences and the challenges of working through misconceptions and partial understandings in the course of learning (Black, Wilson & Yao, 2011). Other challenges concerning healing, pain, and trauma arise in health care contexts (Hibbard et al., 2004; Hobart et al., 2007).

Latour (2004, p. 219) concurs with Wright, saying that:

> ...neither distance nor empathy defines well-articulated science. You may fail to register the counter-questioning of those you interrogate, either because you are too distanced or because you are drowning them in your own empathy. Distance and empathy, to be useful, have to be subservient to this other touchstone: do they help maximize the occasion for the phenomenon at hand to raise its own questions against the original intentions of the investigator—including of course the generous 'empathic' intentions?
>
> It must be clear, according to this formulation, that abstaining from biases and prejudices is a very poor way of handling a protocol. To the contrary, one must have as many prejudices, biases as possible, to put them at risk in the setting and provide occasions of manipulation for the entities to show their mettle. It is not passion, nor theories, nor preconceptions that are in themselves bad, they only become so when they do not provide occasions for the phenomena to differ. (p. 219).

Most measurement practices in psychology and the social sciences systematically ignore questions raised by the phenomena measured against the investigator's intention. The assignment of numbers according to rules is typically conducted with no methodological route provided for the object of investigation to assert itself. By hypothesizing the definition of invariant, additive units of measurement, however, the approach to quantification taken up by Wright puts a priority on qualitative meaning over and above mere numeric manipulations (Fisher, 2004; Mundy, 1986; Maul, 2017). Wright's (1968, 1977, 1999; Wright & Masters, 1982; Wright & Stone, 1979, 1999; Fisher & Wright, 1994; Wilson & Fisher, 2017) focus on construct maps and theories, introduction of individual-level model fit statistics, emphasis on individualized uncertainty ranges, and various data quality evaluations all open up opportunities for measured constructs to assert themselves as independent forces that may contradict the researcher's desires for unfettered control of the research situation, unbothered by the objections of the phenomenon studied.

The general failure in social science to test the hypothesis of a meaningful unit of comparison is a primary means by which the modern ethos prevents researchers from putting their subjectivities at risk relative to the phenomenon measured. The freedom to assign numbers to observations according to a rule allows researchers to do without testing the data obtained to see if a meaningful unit of measurement can be calibrated. This allows them a great deal of latitude in deciding what questions to ask, and in selecting statistical approaches to their data. The end result is a prioritization of the researchers' subjectivities over those of the persons participating in the research, as well as over

the objectivity of the phenomenon measured. The so-called measures produced in this context are then constitutionally incapable of representing a generalized fusion of object and subject horizons accessible on a broad scale across individual instances of things and people.

As described by Wright (1958), understanding begins as a subjective guess that may well be colored by psychological projections, denials, or repressions, but is validated or negated when it encounters the objective text of, for instance, a student's response to a question, a colleague's differing opinion, or an alternative theory. And so it is that Guess and validation are in a sense circularly related as subjective and objective approaches to the text. But this circle is not a vicious one. That would be the case if we were unable to escape the kind of 'self-confirmability' which threatens the relation between guess and validation (Ricoeur 1976, 79).

We are able to escape the vicious circle of self-confirmability to the extent that experimental tests of hypotheses concerning the repeatable reproduction of a unit of measurement support the existence of a persistently identifiable thing in the world. Meaningful patterns are grasped only by means of personal commitments to finding them that are mediated by the explanatory capacity of the text to disclose a world (Ricoeur, 1981, p. 220). Subjective desires for particular meanings are qualified by what the text says, i.e., what it refers to in the world pushes back as the objective counterpart of a particular subjective relation to it.

Occasions for the phenomena to differ as to what it is, as Kuhn (1961) stresses, are a primary function of measurement in science. Scientific instruments are readable technologies calibrated via theory and experiement to measure in a repeatably reproducible unit within a given range of uncertainty. The explanatory value provided by experimental and theoretical proofs is embedded in the semiotic system of measuring instruments such that, when an instrument is read, the subject experiences the projection of a new world horizon. This projection can be as simple as learning that the oven temperature is high enough to put in the bread dough, or as profound as understanding Huck Finn's lesson as to the differences between friendship and compliance with written rules.

As Gadamer (1981, p. 164) put it, "the fruitfulness of scientific questioning is defined in an adequate manner if it is really open to answers in the sense that experience can refuse the anticipated confirmation." Wright's contrast of subjective and objective approaches to learning then takes the form of the "dialectic of belonging and distanciation" described by Ricoeur (1976, p. 79) as a circularly related process of guess and validation. What Wright (1958) accomplished in his formulation of a personal approach to learning is an independent start at what has been called a simultaneously objective and subjective "joint epistemic project addressing the historically changing and mutually conditioning relation of 'inside' and 'outside' knowledge" (Galison, 2008, p. 293). This project, in Ricoeur's (1981) terms, does not imply any kind of direct congeniality of one soul with another. Nothing is less intersubjective or dialogical than the encounter with a text; what Gadamer calls the 'fusion of horizons' expresses the convergence of the world horizons of the writer and the reader. The ideality of the text remains the mediator in this process of the fusion of horizons (pp. 191-192).

In this process, Ricoeur continues,

> what is 'made our own' is not something mental, not the intention of another subject, nor some design supposedly hidden behind the text; rather, it is the projection of a world, the proposal of a mode-of-being-in-the-world, which the text discloses in front of itself by means of its non-ostensive references. Far from saying that a subject, who already masters his own being-in-the-world, projects the a priori of his own understanding and interpolates this a priori in the text, I shall say that appropriation is the process by which the revelation of new modes of being or if you prefer Wittgenstein to Heidegger, new 'forms of life'- gives the subject new capacities for knowing himself. If the reference of a text is the projection of a world, then it is not in the first instance the reader who projects himself. The reader is rather broadened in his capacity to project himself by receiving a new mode of being from the text itself. (pp. 192-193).

The fusion of subject and object horizons opens up a new world of possibilities for the reader of an instrument, be it a book, a smartphone screen, a steam engine, or a thermometer. Instead of a suspicious

delving into subjective depths of hidden meanings separate from the objective text, a pragmatic perspective sees a movement from a meaningful connection of words and things to the actions of an agent influencing events (Fisher & Cavanagh, 2016). This deepens and broadens previous descriptions of measurement as involving a dualism of objective and intersubjective evaluations.

Intersubjectivity in the context of recent philosophical developments in model-based measurement, for instance, concerns the expression of measures, and degrees of belief in them, in the form of "information interpretable in the same way by different users in different places and times" (Mari & Carbone, 2012, p. 2109). This sense of the reliability of interpretable measurement information is left undeveloped in a way that perpetuates the assumption of a dualistic separation of subject and object, wherein a self-sufficient subject takes possession of something mental from someone else or uncovers something hidden within the instrument read. Instead of focusing on objectivity and intersubjectivity, the reliability of measurement information might, then, be characterized in a way opening onto richer possibilities if it were considered in terms of fused subject-object horizons.

Words themselves embody ostensive pointers to objective things in the world, nonostensive references to things not present, and also serve as models of subjective experience. The use of a particular language's words, grammars, dictionaries, vocabularies, semantics, etc. projects subjective expectations as to the listener's or reader's ability to understand. Language use implies a model of the interlocutor that restrictively focuses on others who understand the language and the level of conceptual complexity expressed. In language use, the horizon of the object's existence fuses to some degree with the horizon of the subject's experience. Both objective things themselves and others with their independent subjective experiences can push back against expectations in ways that teach useful lessons.

Learning from our mistakes and creating opportunities for noticing them in the first place are not of value just in education, but are also a part of the playful trial and error processes of science. As Kuhn (1961) emphasized, having rigorously formulated and clear expectations has a proven track record in science as the means by which anomalies are most decisively revealed. Systematically creating new opportunities for revealing the exceptions that prove, in the sense of test, the rule of invariant measurement should, then, become a priority of psychological and social science research policy in the coming years. Broad scale implementations producing a new "prodigious multiplication" of unified subject-object monads like that noted by Tarde (2012) in 1895 will tap network dynamics to support new forms of creativity and innovation in health care, education, and other fields (Fisher, 2012).

3. Conclusion

Experimental reproducibility and theoretical prediction are two of the factors involved in the social accomplishment of the trust packaged in instruments (Wise, 1995, p. 358). Measured constructs initially act as agents provoking agreement among observers as to their independent real existence. Then those agents of agreement are transformed into products of the observers' agreement on technical issues of unit size, uncertainty tolerances, and traceability to standards. Instruments developed in a science that establishes reliable associations between words and things work as advertised to provide end users with otherwise unavailable or difficult to obtain information or products. The test and survey tools of psychology and the social sciences, however, are not typically evaluated or constructed in this way. Research and practice dating back to Thurstone (1926, 1928), Rasch (1960, 1977), Wright (1968, 1977, 1999), and others, however, have demonstrated thousands of times over the viability of instrument calibration and measurement methods capable of extending nondualistic everyday thinking and language into new areas (Andrich, 1988, 2010; Bond & Fox, 2015; Dawson, 2004; Embretson, 2010; Engelhard, 2012; Fischer, 1973; Stenner, et al., 2013; Fisher & Stenner, 2016; Wilson, 2005). This sphere of measurement models real things in the world in ways that focus on hypothetical tests and theoretical predictions capable of producing the social accomplishment of packaged trust.

For instance, over the course of the 1960s and into the 1970s, repeated study of reading comprehension led to the idea that various tests measuring it could be equated to a common unit of comparison. The Anchor Test Study put seven major reading tests on the same scale of measurement in research involving 350,000

students in all 50 U.S. states (Jaeger, 1973). The resulting National Reference Scale for Reading (Rentz & Bashaw, 1977), however, never got off the ground because it was calibrated only from data. The expense of continuously calibrating new items into an enormous bank in an ongoing series of empirical experiments doomed the project. The data-dependency of this approach to measurement short-circuited the economy of language by not connecting the number words on the scale and the real thing in the world being measured with a conceptual determination established as stable and predictable. The connections between the measuring scale, reading comprehension, and text reading difficulty were not sufficiently well elaborated in a theory of the construct to enable a generalized reference function to take hold.

Several years later, however, research demonstrating an explanatory model of visual attention and short term memory (Wright & Stone, 1979; Stenner & Smith, 1982) set the stage for the development of a theory of text complexity enabling efficient predictive control over item difficulty (Stenner, Smith & Burdick, 1983; Stenner, 1996; Stenner & Stone, 2010; Stenner, et al., 2013; Fisher & Stenner, 2016). Now experimental reproducibility was combined with a theory explaining variation in item difficulty in terms of sentence length and word commonality, such that items written from specifications would calibrate at the expected locations on the scale within a small range of uncertainty. This careful alignment and coordination of data, theory, and instrument has resulted in educational publishers creating and widely distributing tools for integrated assessment and instruction by which educators can match individual readers to text with predictable results in reading comprehension. Over 32 million students in the U.S. have at least one measure on this scale per year, with hundreds of thousands of books, and millions of articles, available for use in differentiated instruction (http://www.lexile.com). Research capitalizing on the decades-long availability of this unit of comparison has revealed fascinating global effects of statewide education policy changes in North Carolina (Williamson, 2018).

Thus, in the same way thermodynamics ows more to the steam engine than ever the steam engine owed thermodynamics (Price, 1986, p. 240), so, too, does reading theory owe more to computerized studies of reading tests than ever these tests owed to reading theory. Similar efforts at developing strong theoretical accounts of constructs measured via tests, surveys, and assessments have been underway in other areas for considerable periods of time (Dawson, 2004; Embretson, 2010; Fischer, 1973; Fisher, 2008). Repeated calibrations of both clinician-assigned ratings of physical functioning (activities of daily living and mobility skills), and patient self-reported ratings, for example, have resulted in multiple instances of (a) the same instrument producing statistically identical item calibrations across samples, and (b) different instruments producing highly correlated calibrations of similar items (Fisher, 1997a, 1997b, 1999; Fisher et al., 1995; Fisher, Harvey & Kilgore, 1995; Hobart et al., 2003; Massof, 2005, 2007; Massof & Ahmadian, 2007; Smith & Taylor, 2004). The empirical data and the limited theoretical developments produced to date (Cano et al., 2014; Fisher, 2008; Massof & McDonnell, 2012; Massof et al., 2013; Powers, Fisher & Massof, 2016) suggest strongly that various groups of instruments measuring the same thing could be made metrologically traceable to standard units if enhanced communication and a platform for quality improvement innovations were desired (Cano at al., 2018; Fisher, 1998, 2000, 2005, 2009; Massof, 2008).

Western culture commonly assumes that objectively existing things in the world are self-evident and are transparently absorbed into social life as a matter of course, on the basis of their factuality. Psychological and social measurement are therefore commonly and mistakenly conceived in modern dualist terms as primarily focused on describing data in the absence of explanatory models, predictive theory, and calibrated instrumentation. The history of measurement in psychology and the social sciences makes it plain that the modern myth is an obstacle to managing the complex associations of objectivity and subjectivity that must be worked out before new innovative products will come into language in these areas. Until experimental tests, calibrated instruments, and predictive theory are systematically valued and coordinated in the nondualistic production of technologically embodied subject-object fusions, psychology, the social sciences, and the allied fields of education, health care, management, social services, etc. will continue contributing to the problems alienating humanity from itself, and preventing it from making itself at home in the world, instead of contributing to the solution of these problems. Our challenge, as Gadamer (1991) put it, is

> the systematic problem of philosophy itself: that the part of lived reality that can enter into the concept is always a flattened version--like every projection of a living bodily existence onto a surface. The gain in unambiguous comprehensibility and repeatable certainty is matched by a loss in stimulating multiplicity of meaning. ...all interpretation makes its object univocal and, by providing access to it, necessarily also obstructs access to it (pp. 7-8).

Thus it is the case that even when we focus on stimulating multiplicities of meaning, we inevitably do so via linguistic expressions that can only flatten lived reality from another perspective, since focusing on the multiplicities will require each of them to be expressed with some degree of unambiguous comprehensibility, resulting in a never-ending spiral of deconstructions. As a counterpoint, it is of interest to note that Derrida (2003) then conversely remarks that:

> "people who read me and think I'm playing with or transgressing norms--which I do, of course--usually don't know what I know: that all of this has not only been made possible by but is constantly in contact with very classical, rigorous, demanding discipline in writing, in 'demonstrating,' in rhetoric. ...the fact that I've been trained in and that I am at some level true to this classical teaching is essential. ... When I take liberties, it's always by measuring the distance from the standards I know or that I've been rigorously trained in." (pp. 62-63).

Language embodies qualitative measures of meaning accessible only by means of reference to standards. Though we commonly presume measurement to be primarily quantitative, qualitative meaningfulness is more fundamental to it than quantities are (Fisher, 2003, 2004; Mari et al., 2016). The advancement of scientific measurement today demands that we abandon the modern concern with mere numeric forms of quantification that presume subject-object dualities. We should attend instead to language as the model of how things come into words by distributing qualitatively meaningful measures via portable technologies embodying fused subject-object horizons throughout interconnected multilevel social ecologies. It is no accident, and not a mere desire for an attractive image, that a stylized yin-yang symbol is the trademark logo of the most widely used educational measurement system of this kind (Fisher & Stenner, 2016; Stenner, 1996; Stenner, et al., 2013; Stenner & Stone, 2010).

After quoting another advocate of monads, Spinoza, Ricoeur (1974, pp. 165-166) goes on to note that "a social ethic cannot spring from a system but from a paradox. It aims at two opposed things: human totality and human singularity." A social ethic capable of this integration will emerge only from meta-systematic resolutions of the paradox of inclusively addressing the needs of humanity as a whole while also vigorously personalizing to the maximum relationships that tend to become anonymous and inhuman in the wake of the quest for a shared human identity, world order, and a generalized economy. Given the concepts and tools available to us, we can confidently take up common cause with Ricoeur (1974, p. 166) when, referring to the paradox of opposed human totality and human singularity, he says, "I want both."

References

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

Andrich, D (2010), Sufficiency and conditional estimation of person parameters in the polytomous Rasch model. *Psychometrika,* **75**(2), 292-308.

Arthur, W. B (2014), *Complexity and the economy.* New York: Oxford University Press.

Barney, M., & Fisher W. P., Jr. (2016), Adaptive measurement and assessment. *Annual Review of Organizational Psychology and Organizational Behavior,* **3**, 469-490.

Black, P., Wilson, M., & Yao, S. (2011), Road maps for learning: A guide to the navigation of learning progressions. Measurement: *Interdisciplinary Research & Perspectives,* **9**, 1-52.

Bohr, N (1963). *Essays 1958-1962 on atomic physics and human knowledge.* New York: John Wiley & Sons.

Bond, T., & Fox, C. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences,* 3d edition. New York: Routledge.

Box, G. E. P. (1979), Robustness in the strategy of scientific model building. In R. L. Launer & G. N. Wilkinson (Eds.), *Robustness in statistics* (pp. 201-235). New York: Academic Press, Inc.

Braudel, F. (1992), *Civilization and capitalism, 15th-18th centuries.* Vols. 1-3. (S. Reynolds, Trans.). Berkeley: University of California Press.

Bud, R., & Cozzens, S. E. (Eds). (1992), *SPIE Institutes: Vol. 9. Invisible connections: Instruments, institutions, and science* (R. F. Potter, Ed.). Bellingham, WA: SPIE Optical Engineering Press.

Butler, T. (2016), *Towards a hermeneutic method for interpretive research in information systems. In Enacting Research Methods in Information Systems:* Volume 2 (pp. 11-39). Springer International Publishing.

Campbell, N. R. (1920), *Physics, the elements.* Cambridge, England: Cambridge University Press.

Candea, M. (Ed.) (2010), *Routledge Advances in Sociology.* Vol. 166: The social after Gabriel Tarde: Debates and assessments. London: Routledge.

Cano, S., Mayhew, A., Glanzman, A. M., Krosschell, K. J., Swoboda, K. J., Main, M. et al., (2014), Rasch analysis of clinical outcome measures in spinal muscular atrophy. *Muscle & Nerve,* **49(3)**, 422-430.

Cano, S., Pendrill, L., Barbic, S., & Fisher, W. P., Jr. (2018), Patient-centred outcome metrology for healthcare decision-making. *Journal of Physics:* Conference Series, 1044, 012057.

Cartwright, N. (1983), *How the laws of physics lie.* New York: Oxford University Press.

Chien, T.-W., Linacre, J. M., & Wang, W.-C. (2011), Examining student ability using KIDMAP fit statistics of Rasch analysis in Excel. In Communications in Computer and Information Science: Vol. 201. *Advances in Information Technology and Education* (pp. 578-585). Berlin: Springer Verlag.

Cohen, I. B. (1985), *Revolution in science.* Cambridge, Massachusetts: Harvard University Press.

Dawson, T. L. (2004), Assessing intellectual development: Three approaches, one sequence. *Journal of Adult Development,* **11(2)**, 71-85.

Derrida, J. (2003), Interview on writing. In G. A. Olson & L. Worsham (Eds.), *Critical intellectuals on writing* (pp. 61-69). Albany, New York: State University of New York Press.

Dewey, J. (1925/1981), The later works of John Dewey, Volume 1, 1925-1953: 1925, Experience and Nature (J. A. Boydston, Ed.). *Collected works of John Dewey*. Carbondale, Illinois: University of Southern Illinois Press.

Dewey, J. (2012), *Unmodern philosophy and modern philosophy* (P. Deen, Ed.). Carbondale, Illinois: Southern Illinois University Press.

Einstein, A. (1954), Physics and reality (S. Bargmann, Trans.). In C. Seelig & and others (Eds.), *Ideas and opinions* (pp. 290-323). New York: Bonanza Books (Original work published 1936).

Embretson, S. E. (2010), Measuring psychological constructs: Advances in model-based approaches. Washington, DC: *American Psychological Association.*

Engelhard, G., Jr. (2012), *Invariant measurement: Using Rasch models in the social, behavioral, and health sciences.* New York: Routledge Academic.

Fischer, G. H. (1973), The linear logistic test model as an instrument in educational research. *Acta Psychologica,* **37**, 359-374.

Fisher, W. P., Jr. (1997a), Physical disability construct convergence across instruments: Towards a universal metric. *Journal of Outcome Measurement,* **1(2)**, 87-113.

Fisher, W. P., Jr. (1997b), What scale-free measurement means to health outcomes research. *Physical Medicine & Rehabilitation State of the Art Reviews,* **11(2)**, 357-373.

Fisher, W. P., Jr. (1998), A research program for accountable and patient-centered health status measures. Journal *of Outcome Measurement,* **2(3)**, 222-239.

Fisher, W. P., Jr. (1999), Foundations for health status metrology: The stability of MOS SF-36 PF-10 calibrations across samples. *Journal of the Louisiana State Medical Society,* **151(11)**, 566-578.

Fisher, W. P., Jr. (2000), Objectivity in psychosocial measurement: What, why, how. *Journal of Outcome Measurement,* **4(2)**, 527-563.

Fisher, W. P., Jr. (2003), Mathematics, measurement, metaphor, metaphysics: Parts I & II. *Theory & Psychology,* **13(6)**, 753-828.

Fisher, W. P., Jr. (2004), Meaning and method in the social sciences. Human Studies: *A Journal for Philosophy and the Social Sciences, 27(4)*, 429-54.

Fisher, W. P., Jr. (2005), Daredevil barnstorming to the tipping point: New aspirations for the human sciences. J*ournal of Applied Measurement*, **6(3)**, 173-179.

Fisher, W. P., Jr. (2008, 2-5 March), A predictive theory for the calibration of physical functioning patient survey items. Presented at the *Second Conference on Patient Reported Outcomes Measurement Information Systems,* Bethesda, Maryland: NIH and NIAMS.

Fisher, W. P., Jr. (2009), Invariance and traceability for measures of human, social, and natural capital: *Theory and application. Measurement: Concerning Foundational Concepts of Measurement Special Issue Section,* **42(9)**, 1278-1287.

Fisher, W. P., Jr. (2012), What the world needs now: A bold plan for new standards. *Standards Engineering,* **64(3)**, 1-5.

Fisher, W. P., Jr. (2017), A practical approach to modeling complex adaptive flows in psychology and social science. *Procedia Computer Science,* **114**, 165-174.

Fisher, W. P., Jr. (2018), How beauty teaches us to understand meaning. *Educational Philosophy and Theory, in review.*

Fisher, W. P., Jr., & Cavanagh, R. (2016), Measurement as a medium for communication and social action, I & II. In Q. Zhang & H. H. Yang (Eds.), Pacific Rim Objective Measurement Symposium (PROMS) *2015 Conference Proceedings* (pp. 153-182). Berlin: Springer-Verlag.

Fisher, W. P., Jr., Harvey, R. F., & Kilgore, K. M. (1995), New developments in functional assessment: Probabilistic models for gold standards. *NeuroRehabilitation,* **5(1)**, 3-25.

Fisher, W. P., Jr., Harvey, R. F., Taylor, P., Kilgore, K. M., & Kelly, C. K. (1995), Rehabits: A common language of functional assessment. *Archives of Physical Medicine and Rehabilitation,* **76(2),** 113-122.

Fisher, W. P., Jr., Oon, E. P.-T., & Benson, S. (2018), Applying Design Thinking to systemic problems in educational assessment information management. *Journal of Physics Conference Series,* **1044,** 012012.

Fisher, W. P., Jr., & Stenner, A. J. (2016), Theory-based metrological traceability in education: A reading measurement network. *Measurement,* **92,** 489-496.

Fisher, W. P., Jr., & Stenner, A. J. (2018), Ecologizing vs modernizing in measurement and metrology. *Journal of Physics: Conference Series,* **1044,** 012025.

Fisher, W. P., Jr., & Wilson, M. (2015), Building a productive trading zone in educational assessment research and practice. Pensamiento Educativo: *Revista de Investigacion Educacional Latinoamericana,* **52(2)**, 55-78.

Fisher, W. P., Jr., & Wright, B. D. (Eds.), (1994), Applications of probabilistic conjoint measurement. International *Journal of Educational Research,* **21(6),** 557-664.

Freud, S. (1920), One of the difficulties of psycho-analysis. *International Journal of Psychoanalysis,* **1**, 17-23.

Gadamer, H.-G. (1981), *Studies in Contemporary German Social Thought. Vol. 2: Reason in the age of science* (T. McCarthy, Ed.) (F. G. Lawrence, Trans.). Cambridge: MIT Press.

Gadamer, H.-G. (1989), *Truth and method* (J. Weinsheimer & D. G. Marshall, Trans.) (Rev. ed.). New York: Crossroad (Original work published 1960).

Gadamer, H.-G. (1991), *Plato's dialectical ethics: Phenomenological interpretations relating to the Philebus* (R. M. *Wallace, Trans.).* New Haven, Connecticut: Yale University Press.

Galison, P. (2008), Image of self. In L. Daston (Ed.), *Things that talk: Object lessons from art and science* (pp. 256-294). New York: Zone Books.

Gasché, R. (2014), "A certain walk to follow:" Derrida and the question of method. Epoché: *A Journal for the History of Philosophy, 18(2)*, 525-550.

Gelven, M. (1984), Eros and projection: Plato and Heidegger. In R. W. Shahan & J. N. Mohanty (Eds.), *Thinking about Being: Aspects of Heidegger's thought* (pp. 125-36). Norman, Oklahoma: Oklahoma University Press.

Gilligan, C. (1982), *In a different voice.* Cambridge, MA: Harvard University Press.

Habermas, J. (1995), *Moral consciousness and communicative action.* Cambridge, Massachusetts: MIT Press.

Hankins, T. L., & Silverman, R. J. (1999), *Instruments and the imagination.* Princeton, New Jersey: Princeton University Press.

Haraway, D. J. (1996), Modest witness: Feminist diffractions in science studies. In P. Galison & D. J. Stump (Eds.), *The disunity of science: Boundaries, contexts, and power* (pp. 428-441). Stanford, California: Stanford University Press.

Harding, S. (2008), *Sciences from below: Feminisms, postcolonialities, and modernities.* Durham, NC: Duke University Press.

Hayek, F. A. (1948), *Individualism and economic order*. Chicago: University of Chicago Press.

Heelan, P. A. (1991), Hermeneutical phenomenology and the philosophy of science. In H. J. Silverman (Ed.), *Continental Philosophy: Vol. IV. Gadamer and hermeneutics* (pp. 213-28). New York: Routledge.

Heelan, P. A. (1994), Galileo, Luther, and the hermeneutics of natural science. In T. J. Stapleton (Ed.), *The question of hermeneutics: Essays in honor of Joseph J. Kockelmans* (pp. 363-374). Contributions to Phenomenology, 17. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Heidegger, M. (1962), *Being and time (J. Macquarrie & E. Robinson, Trans.*). New York: Harper & Row (Original work published 1927).

Heidegger, M. (1967), *What is a thing? (W. B. Barton, Jr. & V. Deutsch, Trans.*). South Bend, Indiana: Regnery/Gateway.

Heidegger, M. (1991), *The principle of reason (R. Lilly, Trans.).* Bloomington, Indiana: Indiana University Press.

Hibbard, J., Stockard, J., Mahoney, E., & Tusler, M. (2004), Development of the Patient Activation Measure (PAM): Conceptualizing and measuring activation in patients and consumers. *Health Services Research,* **39**(4, Part I), 1005-1026.

Hobart, J. C., Cano, S. J., O'Connor, R. J., Kinos, S., Heinzlef, O., Roullet, E. P., C. et al. (2003), Multiple Sclerosis Impact Scale-29 (MSIS-29): Measurement stability across eight European countries. *Multiple Sclerosis,* **9**, S23.

Hobart, J. C., Cano, S. J., Zajicek, J. P., & Thompson, A. J. (2007), Rating scales as outcome measures for clinical trials in neurology: Problems, solutions, and recommendations. *Lancet Neurology,* **6**, 1094-1105.

Holton, G. (1988), *Thematic origins of scientific thought: Kepler to Einstein (Revised ed.).* Cambridge, Massachusetts: Harvard University Press.

Husserl, E. (1970), *The crisis of European sciences and transcendental phenomenology: An introduction to phenomenological philosophy (D. Carr, Trans.).* Evanston, Illinois: Northwestern University Press (Original work published 1954).

Hutchins, E. (2010), *Cognitive ecology.* Topics in Cognitive Science, **2**, 705-715.

Hutchins, E. (2014), The cultural ecosystem of human cognition. *Philosophical Psychology,* **27**(1), 34-49.

Huxley, T. H. (1862), *On our knowledge of the causes of the phenomena of organic nature.* London: Robert Hardwicke.

Ihde, D. (1983), The historical and ontological priority of technology over science. *In Existential technics* (pp. 25-46). Albany, New York: State University of New York Press.

Ihde, D. (1990), *Technology and the lifeworld: From garden to earth. Bloomington,* Indiana: Indiana University Press.

Ihde, D. (1991), *Instrumental realism: The interface between philosophy of science and philosophy of technology. The Indiana Series in the Philosophy of Technology).* Bloomington, Indiana: Indiana University Press.

Irigaray, L. (1984), *An ethics of sexual difference.* Ithaca, New York: Cornell University Press.

Jaeger, R. M. (1973), The national test equating study in reading (The Anchor Test Study). *Measurement in Education,* **4**, 1-8.

Kuhn, T. S. (1961), The function of measurement in modern physical science. Isis, 52(168), 161-193. (Rpt. in T. S. Kuhn, (Ed.). (1977), *The essential tension: Selected studies in scientific tradition and change* (pp. 178-224). Chicago: University of Chicago Press).

Kuhn, T. S. (1970), The structure of scientific revolutions. In O. Neurath, R. Carnap & C. Morris (Eds.), *Foundations of the unity of science: Toward an international encyclopedia of unified science* (Vols. II, Nos. 1-9, pp. 53-272). Chicago, Illinois: University of Chicago Press.

Kuhn, T. S. (1977), *The essential tension: Selected studies in scientific tradition and change.* Chicago, Illinois: University of Chicago Press.

Latour, B. (1983), Give me a laboratory and I will move the world. In K. D. Knorr-Cetina & M. Mulkay (Eds.), *Science observed: Perspectives on the social study of science* (pp. 141-170). London, England: Sage Publications.

Latour, B. (1987), *Science in action: How to follow scientists and engineers through society.* New York: Harvard University Press.

Latour, B. (1990), Postmodern? No, simply amodern: Steps towards an anthropology of science. *Studies in History and Philosophy of Science,* **21(1)**, 145-171.

Latour, B. (1991), The impact of science studies on political philosophy. *Science, Technology, & Human Values*, **16(1),** 3-19.

Latour, B. (1993), *We have never been modern.* Cambridge, Massachusetts: Harvard University Press.

Latour, B. (1998), To modernise or ecologise? That is the question. In B. Braun & N. Castree (Eds.), *Remaking reality: Nature at the millennium* (pp. 221-242). London: Routledge.

Latour, B. (2002), Gabriel Tarde and the end of the social. In P. Joyce (Ed.), *The social in question: New bearings* (pp. 117-132). London: Routledge.

Latour, B. (2004), How to talk about the body? The normative dimension of science studies. *Body & Society,* 10(2-3), 205-229.

Latour, B. (2010), Tarde's idea of quantification. In M. Candea (Ed.), *The social after Gabriel Tarde: Debates and assessments* (pp. 145-162). London: Routledge.

Latour, B. (2011), Love your monsters: Why we must care for our technologies as we do our children. *Breakthrough Journal,* **2,** 21-28.

Latour, B. (2013), *An inquiry into modes of existence (C. Porter, Trans.).* Cambridge, Massachusetts: Harvard University Press.

Latour, B., & Lépinay, V. A. (2090), *The science of passionate interests: An introduction to Gabriel Tarde's economic anthropology.* Chicago: Prickly Paradigm Press.

Luce, R. D., & Tukey, J. W. (1964).,Simultaneous conjoint measurement: A new kind of fundamental measurement. *Journal of Mathematical Psychology,* **1(1)**, 1-27.

Lunz, M. E., Bergstrom, B. A., & Gershon, R. C. (1994), Computer adaptive testing. *International Journal of Educational Research,* **21(6)**, 623-634.

Mach, E. (1919), *The science of mechanics: A critical and historical account of its development (T. J. McCormack, Trans.)* (4th ed.). Chicago: The Open Court Publishing Co.

Mari, L., Carbone, P., & Petri, D. (2012), Measurement fundamentals: A pragmatic view. *IEEE Transactions on Instrumentation and Measurement,* **61(8)**, 2107-2115.

Mari, L., Maul, A., Irribara, D. T., & Wilson, M. (2016), Quantities, quantification, and the necessary and sufficient conditions for measurement. *Measurement,* **100**, 115-121.

Mari, L., & Wilson, M. (2014), An introduction to the Rasch measurement approach for metrologists. *Measurement,* **51**, 315-327.

Massof, R. W. (2005), Application of stochastic measurement models to visual function rating scale questionnaires. *Ophthalmic Epidemiology,* 12(2), 103-124.

Massof, R. W. (2007), An interval-scaled scoring algorithm for visual function questionnaires. *Optometry & Vision Science,* **84(8),** E690-E705.

Massof, R. W. (2008), Editorial: Moving toward scientific measurements of quality of life. *Ophthalmic Epidemiology,* **15**, 209-211.

Massof, R. W., & Ahmadian, L. (2007), What do different visual function questionnaires measure? *Ophthalmic Epidemiology,* **14(4)**, 198-204.

Massof, R. W., & McDonnell, P. J. (2012), Latent dry eye disease state variable. *Investigative Ophthalmology & Visual Science,* **53(4),** 1905-1916.

Massof, R. W., Schmidt, K., Laby, D. M., Kirschen, D., & Meadows, D. (2013), Merging psychophysical and psychometric theory to estimate global visual state measures from forced choices. *Journal of Physics Conference Series,* **459(1),** 012027.

Masters, G. N., Adams, R. J., & Lokan, J. (1994), Mapping student achievement. International *Journal of Educational Research,* **21(6)**, 595-610.

Maul, A. (2017), Rethinking traditional methods of survey validation. *Measurement: Interdisciplinary Research & Perspectives*, **15(2),** 1-19.

Maul, A., Mari, L., Torres Irribarra, D., & Wilson, M. (2018), The quality of measurement results in terms of the structural features of the measurement process. *Measurement,* **116,** 611-620.

Mead, R. J. (2009), The ISR: Intelligent Student Reports. *Journal of Applied Measurement,* **10(2),** 208-224.

Michell, J. (1986), Measurement scales and statistics: A clash of paradigms. *Psychological Bulletin,* **100,** 398-407.

Michell, J. (2000), Normal science, pathological science and psychometrics. *Theory & Psychology,* **10(5),** 639-667.

Mohr, M. (2000), Emerging from nonduality: Koan practice in the Rinzai tradition since Hakuin. In S. Heine & D. S. Wright (Eds.) *The Koan: Texts and contexts in Zen Buddhism.* Oxford: Oxford University Press.

Mundy, B. (1986), On the general theory of meaningful representation. *Synthese,* **67(3),** 391-437.

Murdoch, J. (1997), Inhuman/nonhuman/human: Actor-network theory and the prospects for a nondualistic and symmetrical perspective on nature and society. *Environment and Planning D: Society and Space,* **15(6),** 731-756.

Nersessian, N. J. (1996), Child's play. *Philosophy of Science, 63,* 542-546.

Nersessian, N. J. (2002), Maxwell and "the method of physical analogy": Model-based reasoning, generic abstraction, and conceptual change. In D. Malament (Ed.), *Reading natural philosophy: Essays in the history and philosophy of science and mathematics* (pp. 129-166). Lasalle, Illinois: Open Court.

Nersessian, N. J. (2006), Model-based reasoning in distributed cognitive systems. *Philosophy of Science,* **73,** 699-709.

Nersessian, N. J. (2008), *Creating scientific concepts.* Cambridge, Massachusetts: MIT Press.

Nersessian, N. J. (2012), Engineering concepts: The interplay between concept formation and modeling practices in bioengineering sciences. *Mind, Culture, and Activity,* **19,** 222-239.

Nersessian, N. J. (2015), Conceptual innovation on the frontiers of science. In E. Margolis & S. Laurence (Eds.), *The conceptual mind: New directions in the study of concepts* (pp. 455-474). Cambridge, MA: MIT Press.

Nietzsche, F. (1967), *The will to power (W. Kaufmann, Ed.) (W. Kaufmann & R. J. Hollingdale, Trans.).* New York: Vintage.

Nussbaum, M. (2013), *Political emotions: Why love matters for justice.* Cambridge, MA: The Belknap Press of Harvard University Press.

Oudeyer, P.-Y. (2006), *Self-organization in the evolution of speech.* New York: Oxford University Press.

Overton, W. F. (2015), Processes, relations and Relational-Developmental-Systems. In W. F. Overton & P. C. M. Molenaar (Eds.), *Theory and Method. Volume 1 of the Handbook of child psychology and developmental science* (7th Ed.) (pp. 9-62). Hoboken, NJ: Wiley.

Pendrill, L. (2014), Man as a measurement instrument [Special Feature]. NCSLi Measure: *The Journal of Measurement Science,* **9(4),** 22-33.

Pendrill, L., & Fisher, W. P., Jr. (2015), Counting and quantification: Comparing psychometric and metrological perspectives on visual perceptions of number. *Measurement,* **71,** 46-55.

Petersen, A. (1968), *Quantum physics and the philosophical tradition.* Cambridge, Massachusetts: MIT Press.

Pickering, A. (2010), After dualism. In R. E. Lee (Ed.), *Questioning nineteenth-century assumptions about knowledge, III: Dualism* (pp. 49-86). Albany, NY: State University of New York Press.

Platt, J. R. (1966), Social chain reactions. In J. R. Platt, *The step to man* (pp. 39-52). New York: John Wiley & Sons. (Rpt. from J. R. Platt. (1961), Bulletin of the Atomic Scientists: Man and His Habitat, Part II, 17, 365-386.)

Powers, M., Fisher, W. P. J., & Massof, R. W. (2016), Modeling visual symptoms and visual skills to measure functional binocular vision. *Journal of Physics Conference Series,* **772(1),** 012045.Price, D. J. d. S. (1986), Of sealing wax and string. In *Little science, big science--and beyond* (pp. 237-253). New York: Columbia University Press.

Railsback, S. F. (2001), Concepts from complex adaptive systems as a framework for individual-based modelling. Ecological Modelling, 139(1), 47-62.

Rasch, G. (1960), *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.

Rasch, G. (1973/2011), All statistical models are wrong! *Rasch Measurement Transactions,* **24(4),** 1309.

Rasch, G. (1977), On specific objectivity: An attempt at formalizing the request for generality and validity of scientific statements. *Danish Yearbook of Philosophy,* **14,** 58-94.

Rentz, R. R., & Bashaw, W. L. (1977), The National Reference Scale for Reading: An application of the Rasch model. *Journal of Educational Measurement,* **14(2),** 161-179.

Ricoeur, P. (1967), Conclusion: The symbol gives rise to thought. In R. N. Anshen (Ed.), *The symbolism of evil* (pp. 347-357). Boston, Massachusetts: Beacon Press.

Ricoeur, P. (1974), Violence and language. In D. Stewart & J. Bien (Eds.), *Political and social essays by Paul Ricoeur* (pp. 88-101). Athens, Ohio: Ohio University Press.

Ricoeur, P. (1976), *Interpretation theory: Discourse and the surplus of meaning.* Fort Worth, Texas: Texas Christian University Press.

Ricoeur, P. (1981), *Hermeneutics and the human sciences: Essays on language, action and interpretation.* Cambridge, England: Cambridge University Press.

Rowlands, M. (2010), *The new science of the mind: From extended mind to embodied phenomenology.* Cambridge, MA: MIT Press.

Ruddick, S. (1989), *Maternal thinking.* Boston, MA: Beacon Press.

Schaffer, S. (1992), 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, WA: SPIE Optical Engineering Press.

Schubert, C., Sydow, J., & Windeler, A. (2013), The means of managing momentum: Bridging technological paths and organisational fields. *Research Policy,* 42, 1389-1405.

Scott, J. C. (1998), *Seeing like a state: How certain schemes to improve the human condition have failed.* New Haven: Yale University Press.

Searle, J. R. (2010), Why dualism (and materialism) fail to account for consciousness. In R. E. Lee (Ed.), *Questioning nineteenth-century assumptions about knowledge, III: Dualism* (pp. 5-30). Albany, NY: State University of New York Press.

Shapin, S. (1989), *The invisible technician.* American Scientist, **77(6),** 554-563.

Smith, P. H. (2006), *The body of the artisan: Art and experience in the Scientific Revolution.* Chicago, Illinois: University of Chicago Press.

Smith, R. M., & Taylor, P. (2004), Equating rehabilitation outcome scales: Developing common metrics. *Journal of Applied Measurement,* **5(3),** 229-242.

Star, S. L., & Griesemer, J. R. (1989), Institutional ecology, 'translations,' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science,* **19(3),** 387-420.

Star, S. L., & Ruhleder, K. (1996), Steps toward an ecology of infrastructure: Design and access for large information spaces. *Information Systems Research,* **7(1),** 111-134.

Stenner, A. J. (1996), *Measuring reading comprehension with the Lexile Framework.* Durham, North Carolina: MetaMetrics, Inc.

Stenner, A. J., Fisher, W. P., Jr., Stone, M. H., & Burdick, D. S. (2013), Causal Rasch models. Frontiers in Psychology: *Quantitative Psychology and Measurement,* **4(536),** 1-14.

Stenner, A. J., & Smith, M., III. (1982), Testing construct theories. *Perceptual and Motor Skills,* **55,** 415-426.

Stenner, A. J., & Stone, M. (2010), Generally objective measurement of human temperature and reading ability: Some corollaries. *Journal of Applied Measurement,* **11(3),** 244-252.

Sutton, J., Harris, C. B., Keil, P. G., & Barnier, A. J. (2010), The psychology of memory, extended cognition, and socially distributed remembering. *Phenomenology and the Cognitive Sciences,* **9(4),** 521-560.

Tarde, G. (1903a), Inter-psychology, the inter-play of human minds. *International Quarterly,* **7,** 59-84.

Tarde, G. (1903b), *The laws of imitation* (E. C. Parsons, Trans.). New York: Henry Holt and Company.

Tarde, G. (2012), *Monadology and sociology (T. Lorenc, Trans.).* Melbourne, Australia: Re.press (Original work published 1895).

Thoreau, H. D. (1854), *Walden, or life in the woods.* Boston: Ticknor and Fields.

Thurstone, L. L. (1926), The scoring of individual performance. *Journal of Educational Psychology,* **17,** 446-457.

Thurstone, L. L. (1928), Attitudes can be measured. *American Journal of Sociology,* XXXIII, 529-544.

Toulmin, S. E. (1953), *The philosophy of science: An introduction.* New York: Hutchinson's University Library.

Whitehead, A. N. (1911), *An introduction to mathematics.* New York: Henry Holt and Co.

Whitehead, A. N. (1925), *Science and the modern world.* New York: Macmillan.

Williamson, G. (2018), Exploring reading and mathematics growth through psychometric innovations applied to longitudinal data. *Cogent Education,* **5(1464424),** 1-29.

Wilson, M. (2005), *Constructing measures: An item response modeling approach.* Mahwah, New Jersey: Lawrence Erlbaum Associates.

Wilson, M. R. (2011), Some notes on the term: "Wright Map." *Rasch Measurement Transactions,* **25(3),** 1331.

Wilson, M., & Fisher, W. (2017), *Psychological and social measurement: The career and contributions of Benjamin D. Wright.* New York: Springer.

Wise, M. N. (1995), Precision: Agent of unity and product of agreement. Part III--"Today Precision Must Be Commonplace." In M. N. Wise (Ed.), *The values of precision* (pp. 352-361). Princeton, New Jersey: Princeton University Press.

Woolley, A. W., & Fuchs, E. (2011), Collective intelligence in the organization of science. *Organization Science,* **22(5),** 1359-1367.

Wright, B. D. (1958), On behalf of a personal approach to learning. The Elementary School Journal, 58, 365-375. (Rpt. in M. Wilson & W. P. Fisher, Jr., (Eds.). (2017), *Psychological and social measurement: The career and contributions of Benjamin D. Wright* (pp. 221-232). New York: Springer Nature).

Wright, B. D. (1968), Sample-free test calibration and person measurement. In *Proceedings of the 1967 invitational conference on testing problems* (pp. 85-101). Princeton, New Jersey: Educational Testing Service.

Wright, B. D. (1977), Solving measurement problems with the Rasch model. *Journal of Educational Measurement,* **14(2),** 97-116.

Wright, B. D. (1999), 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.

Wright, B. D., Mead, R. J., & Ludlow, L. H. (1980), *KIDMAP: person-by-item interaction mapping* (Tech. Rep. No. MESA Memorandum #29). Chicago: MESA Press [http://www.rasch.org/memo29.pdf].

Wright, B. D., & Stone, M. H. (1979), *Best test design: Rasch measurement*. Chicago, Illinois: MESA Press.

Wright, B. D., & Stone, M. H. (1999), *Measurement essentials.* Wilmington, DE: Wide Range, Inc. [http://www.rasch.org/measess/me-all.pdf].

Yandell, K. E. (1994), *The epistemology of religious experience.* Cambridge, England: Cambridge University Press.

Author Profiles:

William P. Fisher, Jr., Ph.D. received his doctorate from the University of Chicago, where he was supported by a Spencer Foundation Dissertation Research Fellowship. He is currently Research Associate in the Graduate School of Education at the University of California, Berkeley. Dr. Fisher is recognized for contributions to measurement theory and practice that span the full range from the philosophical to the applied across a diverse range of fields.

Magnetic Quantities: Healthcare Sector Measuring Demands and International Infrastructure for providing Metrological Traceability

Elisabeth Costa Monteiro
betth@puc-rio.br
Pontifical Catholic University of Rio de Janeiro

Abstract:

Innovations in the healthcare sector along the last half-century have created a growing demand for metrological traceability of measurement results regarding magnetic quantities. This work discusses these challenging requests for traceability, from the lowest to the highest values of magnetic flux densities associated with the recently developed biomedical technologies as well as other relevant health demands. The worldwide availability of quality metrological infrastructure for providing measurement standards appropriate to the demanding quantity range is analyzed based on the information of the BIPM key comparison database (KCDB) appendices. Results indicate that the demands for metrologically traceable measurement results regarding power frequency magnetic field exposure have demonstrated to be a notable driver toward the establishing of metrological infrastructure by National Metrology Institutes. Although limitations of available measurement standards for magnetic quantities produce significant impacts on the possibility of meeting medical innovations' demands, the recent fast expansion of comprehensiveness of declared CMC capabilities for ensuring metrological traceability of magnetic quantities points toward the possibility of meeting the demanding values from healthcare sector in the near future.

Keywords: healthcare technologies, magnetic quantities, metrological traceability

1. Introduction

Reliability of measurement results, as well as its worldwide comparability, is assured with the possibility to obtain metrological traceability to the International System of Units (SI). Traceable calibrations are required to ensure appropriate function of measuring instruments used in production and testing processes of a wide variety of industries; safety and performance of equipment used in healthcare; environmental safety; and Research & Development sector. In 1927, driven by the expansion of the range of scientific domains demanding the International Committee for Weights and Measures (CIPM) attention under the Meter Convention, signed on May 20, 1875, the first CIPM Consultative Committee was created, the Consultative Committee for Electricity (CCE). However, it was only seventy years later, in 1997, that the full denomination of the committee was given as Consultative Committee for Electricity and Magnetism (CCEM). The custodians of the national measurement standards necessary to support the metrological traceability chain to SI system in a country are the National Metrology Institutes (NMIs). By June 2018, Calibration and Measurement Capabilities (CMC) declarations for magnetic quantities, namely magnetic fields at frequencies below 50 kHz, were included in Appendix C of the BIPM key comparison database (KCDB) by 18 NMIs, being most of the measurands related to low-frequency magnetic flux density measurements. In turn, the demands for traceable calibration chain between magnetic measurements and the SI is growing all over the world and so is the number of countries with CMC declared for these measurands. Among the challenging demands to be met by standardization associated with magnetic quantities are those arisen from the high rate of innovation of the healthcare sector. However, it is particularly recent, in 1999, when the 21st General Conference on Weights and Measures (CGPM) explicitly considered, in its resolutions, the demands for metrological traceability of measurement results derived from biomedical sectors (Costa Monteiro, 2007; Costa Monteiro and Leon, 2015; Costa Monteiro, 2017). The Consultative Committee for Amount of Substance: Metrology in Chemistry (CCQM), which was created in 1993, took on the task of promoting the development of infrastructure to provide metrological traceability in the health sector, but naturally restricted efforts toward biochemical, biomolecular, and biotechnological aspects (Costa Monteiro, 2007; Costa Monteiro & Leon, 2015; Costa Monteiro 2017). Along the last half-century, several innovations using physical principles based on magnetic quantities and aimed at providing non-invasive diagnosis and treatment are being introduced in the clinical environment. These recently available biomedical tools create new demands for metrological traceability of measurement results. In the present work, these emerging magnetic measuring requests for traceability to SI originating from the health sector are described and discussed along with the possibility of being attended by measurement standards provided by the available worldwide quality infrastructure.

2. Health technologies based on magnetic quantities demanding metrological traceability

In the context of the recent progress in the healthcare sector of the last decades, the early seventies have witnessed the birth of biomedical technologies providing innovations in the non-invasive diagnosis and treatment. These new developments making use of operating principles & based on magnetic quantities generated a growing request for metrological traceability of magnetic measurement results for a variety of intensity and frequency ranges. This section presents some examples of these demands emerging from medical environment.

2.1 Biomagnetic field sensors

Since 1970, with the development of highly sensitive magnetometers based on superconductivity, the Superconducting Quantum Interference Device (SQUID), scientific studies of the ultra-weak biomagnetic signal were introduced (Zimmerman et al., 1970). The magnetic flux densities generated by biomagnetic sources comprise values between femtotesla and nanotesla, and their frequencies are in the range of DC to 1 kHz (Andrä &Nowak, 2007).

Biomagnetic fields are generated by the bioelectrical activity of excitable tissues or by magnetic particles placed in the organism. The non-invasive, contactless, and harmless detection of the biomagnetic field allows to obtain information about the temporal and spatial distribution of the field sources, making it possible to identify the location and propagation of bioelectrical activity in excitable tissues, the so-called primary current sources, without being disturbed by the non-homogeneity of the torso volume conductor (Costa Monteiro et al, 1987; Andrä & Nowak, 2007). This information cannot be reached by the conventional measurement of the electric potential difference using surface electrodes, as in Electrocardiography (ECG), Electroencephalography (EEG), Electromyography (EMG), among others, which, being based on the currents generated in the volume conductor, the secondary currents, these measurements are strongly influenced by the non-homogeneity of conductivities of body tissues (Costa Monteiro et al, 1987; Costa Monteiro et al, 2001a). The biomagnetic field measurement allows, for instance, a non-invasive access to the propagation of bioelectric activity in cardiac tissue, which consists of primary current, being undetectable by surface ECG, unless biopotential electrodes are placed directly on the heart muscle (Costa Monteiro et al, 2001a; Costa Monteiro et al, 2001b; Yamada & Yamaguchi 2005) and precise localization of foreign bodies for surgical removal, significantly reducing the procedure time, the exposure to radiation during intervention, and odds of failure, ensuring successful treatment outcomes (Costa Monteiro et al, 2000).

Considering the SQUID drawbacks of needing cryogenic cooling, its high cost of fabrication, installation and operation, alternative sensors have been studied as a new possibility for measuring the ultra-weak biomagnetic fields, based on Giant Magnetoimpedance (GMI) (Cavalcanti et al, 2008; Silva et al, 2011) showing the most promising results. Recent studies indicate that transducers based on the impedance phase characteristics of GMI sensing elements are considerably more sensitive than the usualy applied magnitude-based GMI magnetic transducers (Silva et al, 2014; Benavides et al, 2018). The development of this high sensitivity phase-based GMI sensor allows the measurement of ultra-weak biomagnetic fields to be pursued by using a low-cost device, which would facilitate the application and dissemination of this non-invasive and innocuous diagnostic tool in clinical environments.

2.2 Magnetic resonance imaging

Magnetic resonance imaging (MRI) uses the physics principle of nuclear magnetic resonance (NMR). The technique allows obtaining chemical and physical information about molecules in multiple planes, enabling clinicians to identify abnormalities in various regions of the organism, with the distinct advantage of better image contrast of the soft tissue in body structures (Andrä & Nowak 2007).

The first scan on a human using nuclear magnetic resonance (NMR), performed in 1971, indicated its potential to detect diseases. Subsequently, clinical Investigations with magnetic resonance imaging (MRI) equipment continuously grew up, introducing, in 1980, the first commercial MRI scanner. In 1992, the functional magnetic resonance imaging (fMRI) expanded the clinical use, allowing identifying brain regions and their functions (Andrä & Nowak, 2007; Roth et al, 2018).

Magnetic resonance operates using the interaction between an external magnetic field, radio frequency energy and nuclei possessing spin. Superconductive magnet immersed in liquid helium (-269 ºC) creates magnetic fields from 0.5 to 3.0 T. Protons generate a nuclear magnetic resonance frequency of 42.577 MHz in a magnetic flux density of 1 T. With the aid of an RF field in the MHz range, and a variable static magnetic field along the longitudinal axis, the sharp resonance absorption of magnetic nuclei in biological tissue is used to obtain the spatial distribution of the nuclear magnetization. In particular, hydrogen atoms, which occur naturally in large numbers, allow medically meaningful images to be produced (Webster, 2010). Future perspectives on improving clinical diagnostic applications, even more, preview the implementation of ultra-high-field MR systems, emitting magnetic flux density intensities up to seven teslas (Roth et al, 2018).

2.3 Transcranial Magnetic Stimulation

The first TMS procedures were introduced over the latter half of the 1980s. The technique consists of applying intense pulses, with magnetic flux densities of the order of two to three teslas and short duration (few kHz). The rapidly changing magnetic field generated by an external coil induces electrical currents sufficient to stimulate bioelectrical activities located in different levels of the brain structure, useful in the treatment of neurological and psychiatric disorders. The induced electric field in the conducting brain tissue is associated with the geometry of the stimulating coil employed. TMS equipment has been continuously evolving by developing several coil configurations suitable for direct stimulation of brain regions with different sizes and depths. Currently, there are several TMS coil designs available: ring coil, the figure-of-eight coil, angulated figure-of-eight coil, double cone, Halo (H) coil, and Halo-circular assembly (HCA) coil. With the advantage of inducing currents in brain regions without stimulating the skin, and therefore being painless, TMS has been used in the treatment of several neuropsychiatric disorders, being useful in treating drug-resistant conditions (Bersani et al, 2013; Palatnik de Souza & Costa Monteiro, 2015; Lu & Ueno, 2017; Palatnik de Souza et al, 2018).

2.4 Instrumentation for monitoring levels of environment exposure to magnetic fields

Considering the risks of electromagnetic fields to health, World Health Organization carried out initiatives towards limiting human exposures since 1989 (Tourab & Babouri, 2015). Founded in 1992, the International Commission on Non-Ionizing Radiation Protection (ICNIRP) develops scientific research and publishes guidelines regarding adverse effects of non-ionizing radiation in human health and the environment, including static magnetic fields and electromagnetic fields. As for 1998, ICNIRP publishes guidelines for limiting exposure to electric, magnetic and electromagnetic fields, defining acceptable reference levels. In the same year, the International Electrotechnical Commission issued the IEC 61786, with requirements for instruments and guidance for measurements of low-frequency magnetic and electric fields concerning exposure of human beings. Subsequently, ICNIRP and IEC publications of guidelines and technical standards, respectively, introduced a significant demand for calibrated instruments for measuring magnetic quantities. For example, regarding the power frequency magnetic field exposure (50/60 Hz), according to the ICNIRP's guideline published in 2010, the magnetic flux density of 200 mT is reference level for general public exposure.

3. NMI measurement capabilities for magnetic quantities compared to health technologies demands

With the advent of the mutual recognition arrangement (CIPM MRA) the equivalence of national measurement standards is established through BIPM intercomparisons. Information on CIPM and RMO key and supplementary comparisons, available in Appendix B of KCDB, by June 2018, indicates a total of 235 intercomparisons carried out by the CCEM, being only twelve of them associated with magnetic quantities (Table 1), namely magnetic fields at frequencies below 50 kHz. More than half of CCEM comparisons consisted of key comparisons (128), being, however, only two of them directed to magnetic quantities. Both the magnetic key comparisons already accomplished were performed to low-frequency magnetic flux density, being concluded in 2000 and 2003, with eight and ten participating NMIs, respectively (Table 1).

Table 1 – Intercomparisons regarding magnetic quantities conducted up to june 2018.

year	RMO/CCEM INTERCOMPARISONS	Magnetic Quantity	Participating NMIs
1998	EUROMET.EM- S1	Comparison of measurements on electrical sheet steels	2
1998 - 2000	EUROMET.EM.M- K1	Comparison of low frequency magnetic flux density	8
1998 - 2001	EUROMET.EM- S13	Comparison of measurements on electric sheet steels	3
2001 - 2002	EUROMET.EM.M- S1	S1 Measurement of magnetic flux	5
2001 -2003	CCEM.M- K1	Comparison of low frequency magnetic flux density	10
2009	COOMET.EM- S9	Comparison of measurements on electrical sheets	4
2010	APMP.EM- S9	Comparison of magnetic flux density standards	2
2011	COOMET.EM- S12	Magnetic losses in isotropic and anisotropic electrical steel	3
2012	APMP.EM- S13	DC magnetic flux density	2
2013 - 2014	APMP.EM- S14	Earth-level DC magnetic flux density	6
2013 - 2015	COOMET.EM- S16	Pulsed electric and magnetic fields	2
2014 - 2015	EURAMET.EM.M- S2	Polarization and specific total power loss in soft magnetic materials	5

In June 2018, among the 103 institutes signatories of CIPM MRA, there were eighteen with internationally recognized CMC for Magnetic fields at frequencies below 50 kHz. Regarding magnetic flux density and applied magnetic field strength, a particularly relevant magnetic quantity for health technologies. Seventeen NMIs had their quality infrastructure declared for DC values, and 13 NMIs for AC values associated with magnetic flux density and applied magnetic field strength. Since 2013, metrological traceability for higher levels of magnetic flux densities is offered. Two NMIs approved their declarations up to 1 or 1.2 T in 2013. Along the following years, other eleven NMIs declared capabilities higher than 1 T. It is worth mentioning that a level of 2.5 T was provided by KRISS (Korea Research Institute of Standards and Science), in 2014, and the highest value achieved, 3.5 T, was declared in 2017, by the Czech Metrology Institute. However, a single CMC declaration, from the Russian Federation NMI, comprises AC magnetic flux densities up to 1 T, within the frequency range of 20 Hz to 200 Hz.

Approved in 2013, magnetic flux density of 0.01 μT is the lowest value for which metrological traceability is being provided. Levels of μT have been declared, since then, by five NMIs for static fields, being three of them approved in 2016 and the fifth one, in 2017; and for AC fields by 13 national institutes.

Comparing the capabilities published in the KCDB for magnetic quantities with the health demands highlighted in section two, it can be noticed a progressive improvement of comprehensiveness of quantities' ranges being provided toward meeting the demanding values of the health sector. The high levels of DC magnetic flux densities up to 3.5 T, recently published in Appendix C evolve to complying with MRI needs. Although achieving magnitudes that could be appropriate to TMS demands, the current CMC does not meet the necessary frequency range, since the equipment emits pulses with around 3 kHz.

Metrological traceability for biomagnetic measurements, in turn, cannot be provided by any of the CMC published, once the demands regarding intensities are lower than nanotesla while the lowermost value declared is 0.01 μT.

Regarding the needs originating from the exposure limits established by ICNIRP guideline for magnetic flux density of 200 mT at power frequency (50/60 Hz), a worldwide concern, the entire group of 13 NMIs with AC magnetic flux density CMC, has declared capabilities complying with the frequency and intensities to provide metrological traceability in the necessary range. The Swedish CMC, however, is restricted to 50 Hz, which does not meet power frequency typically used in the American Continent and parts of Asia; and covers 0.1 μT to 300 μT, which is a narrow range around the ICNIRP limit.

Human safety regarding exposure limits for magnetic flux densities has been widely addressed by international technical standards and national regulatory documents, resulting in a triggering factor for the NMIs introduction of infrastructure to provide traceability of magnetic quantities. The IEC 61786 series, published since 1998 by IEC TC 108, deals with the measurement of low-frequency magnetic and electric fields concerning human exposure. Later, additional documents were produced such as the IEC 62311:2007 for assessment of electronic and electrical equipment related to exposure restrictions, as well as, the IEC 62110:2009 addressing measurement procedures with regard to public exposure for electric and magnetic field levels generated by AC power systems. The regulatory effect can be exemplified with the Brazilian NMI, the eighteenth among the set of 18 institutes to declare CMC for magnetic quantities, with approval received on 27 March 2017. Following the publication, in 2009, of a Federal Law 11.934 setting maximum limits for human exposure, the Brazilian Electrical Energy Regulatory Agency (ANEEL) issued normative resolutions establishing the requirements of the Federal Law, with restrictions for time-varying electric and magnetic fields with frequencies of 60 Hz, the power frequency in Brazil (França et al. 2012, Mendonça 2013), according to the limits established by ICNIRP guidelines, in 2010. The regulatory document required using measuring instruments calibrated by accredited laboratories, providing the appropriate evidence of metrological traceability. As a consequence, the Brazilian NMI invested efforts to implement the infrastructure for ensuring national traceability for measurements within the requested range and, since March 2017, it declared CMC for low-frequency magnetic flux density from 1.1 μT to 700 μT, 50/60 Hz; accrediting, on May 2017, the first calibration laboratory for magnetic quantities.

Discussion and Conclusion

Along the last five decades, important innovations in the non-invasive diagnosis and treatment have generated a growing demand for metrological traceability of measurement results regarding magnetic quantities. Research and development of biomedical technologies created requests from ultra-low magnetic flux densities from femtotesla to nanotesla levels (DC to kHz) for magnetic measurements allowing the non-invasive, contactless, and innocuous assess of biomagnetic sources; up to ultra-high values of few teslas, associated with technologies for transcranial magnetic stimulation and magnetic resonance imaging. Emphasized by WHO since 1989, followed by the publication of several ICNIRP guidelines with exposure limits, a series of IEC international technical standards and national regulatory documents; a worldwide concern regarding the risks of human exposure to magnetic fields has been, in turn, a decisive demand for the introduction of infrastructure to provide traceability of magnetic quantities.

High levels of DC magnetic flux densities, up to 3.5 T, of recent declarations published in the Appendix B of KCDB shows that only 5 % of CCEM key and supplementary comparisons are associated with magnetic fields at frequencies below 50 kHz as well as, Appendix C indicates that only 17 % of the institutes signatories of CIPM MRA (18 NMIs) possess internationally recognized CMC for these magnetic quantities. Appendix C of KCDB, drifts toward complying with MRI needs. Although achieving magnitudes that could be appropriate to TMS demands, the current CMC does not meet the necessary frequency range, since the equipment emits pulses with around 3 kHz.

While appropriate measurement standards for demands associated with the mentioned biomedical technologies of high clinical relevance are not fully available, the need of metrologically traceable measurement results regarding power frequency magnetic field exposure (50/60 Hz) has been the primary driver toward the establishing of metrological infrastructure, being provided by the whole set of NMIs with CMC for AC magnetic flux densities.

Although the highlighted limitations, the recent improvement of the international infrastructure for ensuring metrological traceability of magnetic quantities, with the expanding comprehensiveness of the declared quantity's ranges, points toward the possibility of reaching the requested values of healthcare sector demands in the near future.

References

Andrä, W. and Nowak, H. (2007), *Magnetism in Medicine: A Handbook.* 2nd Ed. WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.

Benavides, L.S., Silva, E.C., Costa Monteiro, E., Barbosa, C.R.H. (2018), High Sensitivity Pressure Transducer based on the phase characteristics of GMI magnetic sensors. *Meas. Sci. Technol.* 29: 035106.

Bersani, F.S., Minichino, A., Enticott, P.G., Mazzarini, L., Khan, N., Antonacci, G., Raccah, R.N., Salviati, M., Delle Chiaie, R., Bersani, G., Fitzgerald, P.B. and Biondi, M. (2013), Deep transcranial magnetic stimulation as a treatment for psychiatric disorders: A comprehensive review. *European Psychiatry* **28:** 30–39.

Cavalcanti, F.M.P., Gusmão, L.A.P., Barbosa, C.R.H., Costa Monteiro, E., Gonçalves, L.A.P. and Machado, F.L.A. (2008), Ring shaped magnetic field transducer based on the GMI effect, *Measurement Science & Technology* **19:** (2008) 1-10.

Costa Monteiro, E., Bruno, A.C., Louro, S.R.W., Costa, A.F. and Ribeiro, P.C. (1987), Magnetic Localization of a Current Dipole Implanted in Dogs. *Phys. in Med. and Biol* **32:** 77-86.

Costa Monteiro, E., Barbosa, C.R.H., Lima, E.A., Ribeiro, P.C., Boechat, P. (2000), Locating steel needles in the human body using a SQUID magnetometer. *Physics in Medicine and Biology* **45:** 8, 2389-2402.

Costa Monteiro, E., Eiselt, M., Giessler, F., Haueisen, J., Barbosa, C.R.H. and Lima, E.A. (2001a), Animal Experimentation Study of Atrial Activity Propagation Using a Multi-channel SQUID *System Biomedizinische Technik.* **46:** 2 70–72.

Costa Monteiro, E.; Schleussner, E., Kausch, S., Grimm, B., Schneider, A., Barbosa, C.R.H., Haueisen, J. (2001b), Fetal cardiac activity analysis during twin pregnancy using a multi-channel SQUID system. *Physica. C, Superconductivity.* **354:** 1 87-90.

Costa Monteiro, E. (2007), Biometrologia: Confiabilidade nas Biomedições e Repercussões Éticas. *Metrologia & Instrumentação* **6:** 6-12.

Costa Monteiro, E. and Leon, L.F. (2015), Metrological Reliability of Medical Devices. *Journal of Physics. Conference Series* **588:** 012032.

Costa Monteiro, E. (2017), Bridging the Boundaries Between Sciences to Overcome Measurement Challenges. *Measurement (Mahwah Print)* **15:** 1-3.

França, A.M., Costa, D.G., Pereira, L.N., Soares, M.A., Vitorio, P.C.O., Nascimento, P.R.M. (2012), Low Frequency Magnetic Flux Density Meter Calibration. In: *Proceedings of the XX IMEKO World Congress,* Busan, Republic of Korea, 9-14 September 2012. pp. 1-3.

Lu, M, Ueno S (2017), Comparison of the induced fields using different coil configurations during deep transcranial magnetic stimulation. *PLoS ONE* **12 (6):** e0178422.

Mendonça, A.F.M. (2013), *Brazilian demand for calibration of measuring instruments for magnetic quantities.,* 2013. MSc. Dissertation, Posgraduate Programme in Metrology. Pontifical Catholic University of Rio de Janeiro. Rio de Janeiro. 106p.

Roth, C.G., Marzio, D.H.D., Guglielmo, F.F. (2018), Contributions of Magnetic Resonance Imaging to Gastroenterological Practice: *MRIs for GIs.* Dig Dis Sci **63:** 1102.

Palatnik de Sousa, I. and Costa Monteiro, E. (2015), Transcranial Magnetic Stimulation Conformity Assessment. In: *XXI IMEKO World Congress: Measurement in Research and Industry* Prague. Czech Republic. 30 August – 4 September 2015. 1 pp 1349-1352.

Palatnik de Souza, I., Barbosa, C.R.H., Costa Monteiro, E. (2018), Safe exposure distances for transcranial magnetic stimulation based on computer simulations. *PeerJ* In press.

Silva, E.C., Gusmão, L.A.P, Barbosa, C.R.H., Costa Monteiro, E. and Machado, F.L.A. (2011), High sensitivity giant magnetoimpedance (GMI) magnetic transducer: magnitude versus phase sensing. *Measurement Science & Technology* 22 035204.

Silva, E.C., Barbosa, C. R. H., Gusmão, LA.P., Leipner Y., Fortaleza, L.G.S., and Costa Monteiro, E. (2014), Point matching: A new electronic method for homogenizing the phase characteristics of giant magnetoimpedance sensors. *Review of Scientific Instruments* **85:** 084708.

Tourab, W., and Babouri, A. (2016), Measurement and Modeling of Personal Exposure to the Electric and Magnetic Fields in the Vicinity of High Voltage Power Lines. *Safety and Health at Work* **7:** 102-110.

Webster, J.G. (2010), *Medical Instrumentation: Application and Design.* 4th Ed. John Wiley & Sons, Hoboken, NJ.

Yamada, S., and Yamaguchi, I. (2005), Magnetocardiograms in Clinical Medicine: Unique Information on Cardiac Ischemia, Arrhythmias, and Fetal Diagnosis. *Intern. Med.* **44 (1):** 1–19, 2005.

Zimmerman, J.E., Thiene, P., and Harding, J.T. (1970), Design and operation of stable rf-biased superconducting point-contact quantum devices, and a note on the properties of perfectly clean metal contacts. *J. Appl. Phys* 42: 1572–1580.

Author Profile:

Elisabeth Costa Monteiro received the M.D. degree from the Federal University of Rio de Janeiro (UFRJ), Brazil, in 1983, and specialized in internal medicine in 1984. She received the M.Sc. and Ph.D. degrees in biophysics from the Institute of Biophysics Carlos Chagas Filho of the UFRJ, in 1988 and 1992, respectively; Postdoctoral Research Fellow at the Institute of Advanced Biomedical Technologies (ITAB) of the Gabriele D'Annunzio University, Italy, from 1992 to 1993. Elisabeth is Full Professor of the Postgraduate Programme in Metrology at the Pontifical Catholic University of Rio de Janeiro (PUC-Rio). Her research interests include development of biomedical instrumentation; bioelectric and biomagnetic field measurements and applications in clinical diagnosis and therapy; contributions of metrology for ensuring reliability of biomedical devices.

Infusion Pumps Calibration Methods

Elsa Batista
ebatista@ipq.pt
IPQ-Instituto Português da Qualidade

Maria do Céu Ferreira
mcferreira@ipq.pt
IPQ-Instituto Português da Qualidade

Abstract:

Nowadays, several types of infusion pumps are commonly used for drug delivery, such as syringe pumps and volumetric pumps. These instruments present different measuring features and capacities according to their use and therapeutic application.
In order to ensure the metrological traceability of these flow and volume measuring equipment it is necessary to use suitable calibration methods and standards.
Two different calibration methods can be used to determine the flow error of infusion pumps. One is the gravimetric method, considered as a primary method and commonly used by the National Metrology Institutes. The other calibration method is a secondary method, which relies on an Infusion Device Analyser (IDA) and is typically used by the hospitals maintenance offices.
The gravimetric method is the most accurate one but the comparison method allows calibration of infusion pumps within the hospital facilities. In this work the two calibration methods are described in detail.
Also the two international research projects that allowed the development of these methods at the Volume and Flow Laboratory of the Portuguese institute for quality are explained.

Keywords: Flow Rate, Infusion pumps, Calibration, Traceability

1.Introduction

Medical infusion instruments are widely used, as they are fundamental for primary health care, namely for providing drugs, nutrition and hydration to patients. Infusion pumps are electronic medical instruments widely used in adult, paediatric and neonatal patients, with the purpose of delivering fluids in a intermittently or continuously manner. These instruments can be used in clinical environments or at home. An infusion pump is normally constituted by a fluid reservoir, a generating flow device, than can also have flow regulation functions and of a set of accessories (lines, administration sets, filters, etc.) that allow the fluid to be transported from the reservoir to the patient.

According to IEC 60601-2-24, infusion pumps can be volumetric or with syringe. Volumetric infusion pumps are used to administer medications and nutrients or food (enteric and parenterel) and can be used in a clinical environment or at home. This type of instrument has a mechanical trigger causes the liquid within the tube to move by peristaltic action, thus enabling the administration of medications and nutrients or food. Syringe pumps presuppose a system of a continuous, individual, resistant tube and without connexions in Y. The pump system is composed by and external electronic infusion pump. A pump with a syringe system is made up of an external electronic infusion pump, an infusion line and a syringe compatible with the system usually disposable. More details concerning the characteristics of different types of pumps can be found in the guide – Metrology in health – Best practices guide (SCH, 2018).

2. Calibration, Error and Traceability

Similar to what happens with any other measurement in a clinical context, the measurement of flow requires the necessary accuracy, as well as the guarantee of conformity to the metrological requirements of the measuring instrument. Amongst other sources, the error and the uncertainty associated with the measurement of flow obtained during calibration of the instrument depend on the conditions of the infusion pump, the type of components/consumables used and the chosen calibration method. The calibration of an instrument also allows ensuring is traceability to the International System of Units.

The metrological definitions of calibration, error, traceability and measurement uncertainty can be found in the International Vocabulary of Metrology (IPQ, 2012).

The determination of the error and uncertainty of a measuring instrument makes it possible to determine and characterize its suitability according to a certain criterion of acceptance stipulated by the manufacturer, by the user or by reference standards, normally defined as Maximum Permissible Error (MPE). Thus, for a given instrument to be accepted and considered for use after calibration or testing, the sum of the absolute value of the error and the uncertainty shall be less than or equal to the absolute value of the MPE.

Due to the fact that there is a vast quantity of infusion pumps in health facilities, priorities are frequently defined according to the criticality of use of the various instruments. The health services should define a calibration plan for the infusion pumps. In addition, it should also be defined the calibration periodicity according to the history of the instrument, the context in which it is used and the manufacturer’s recommendations.

In accordance with the performance of the instrument and of the agreed acceptance criteria, the periodicity of the initially calibration defined can be changed as long as it is properly justified.

3. International and national projects regarding infusion pumps

EURAMET – European association of national metrology laboratories started in 2007 the European Metrology Research Program (EMRP).

This research program allowed collaboration between the National Metrology Laboratories (NMI), Universities and industry, trough join research programs in several strategic areas, one of those is health.

In 2007 it was identified by several NMIs that the perfusion technology had underestimated risks namely:

- Difficulty associated with micro flow measurement and control (<1 mL/h);
- Difficulty measuring and controlling the final concentration using multiple pumps;
- The characteristics of the infusion systems are not fully understood, in particular as regards to delay in dosing, compliance, flow stability and impact of variation of normal operating conditions;
- Metrological traceability was not assured for values lower than 0.5 mL/h, because the metrological infrastructure was not fully developed;
- No validation of measurements for flow values below 100 mL/h;
- There are no common usage protocols that include the entire system and accessories.

The Metrology for Drug Delivery (MeDD) project funded by the European Metrology Research programme (EMRP), that was developed between 2012-2015, had as main focus such calibration methods. In this joint research programme (JRP), several Metrology Institutes (Swiss Federal Institute of Metrology - METAS, Danish Technological Institute - DTI, Centre Technique des Industries Aérauliques et Thermiques, France - CETIAT, Portuguese Institute for Quality - IPQ and Dutch Metrology Institute - VSL) developed the primary standards for liquid flow rate (Batista et al., 2015).
The outcomes of this project were discussed in several international conferences and presented in scientific papers, reports and best practice guides that can be found in www.drugmetrology.com.

This project allowed the Volume and Flow Laboratory of IPQ to develop gravimetric standards for flow measurements in the range of 0.12 mL/h to 600 mL/h, with associated uncertainties of 0.3 % and 2.5 %, resulting in new published CMCs in the BIPM - Bureau International des Poids et Mesures, for infusion pumps, flow meters and flow analyzers.

Aiming to disseminate the knowledge obtained from MeDD JRP, a new project (Support for Impact Project (SIP) 15SIP03 – Infusion Uptake) funded by the European Metrology of Innovation and Research Programme (EMPIR) was started in May of 2016 and will end in 2019. This project had the participation of 4 NMIs, including IPQ and seven national hospitals. The 15SIP03 JRP has two main goals:

a) To develop an E-learning module made available on the E-learning platform of the European Society for Intensive Care Medicine (ESICM), with the aim to create awareness and understanding of multi infusion risks and thereby reducing dosing errors and increasing the quality of medical treatment.

b) To incorporaWWte the best metrology practices relating calibration of infusion devices in ISO standards, namely ISO 7886-2 and IEC 60601-2-24.

4. Calibration of infusion pumps

The flow calibration of infusion systems can be carried out through the standard gravimetric method ISO 60601-2-24:2012 and ISO 7886:1996 or by using a Flow Infusion Pump Analyser (IDA).

4.1. Gravimetric method

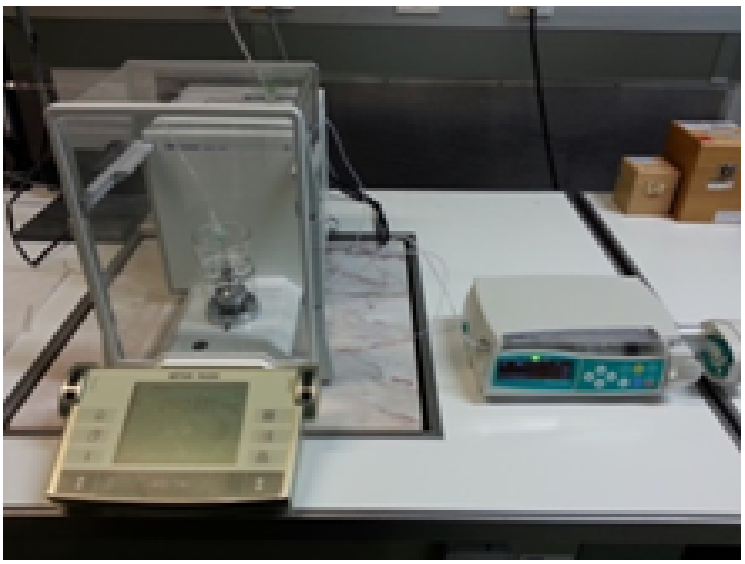
The gravimetric method is considered as a primary method and is commonly used by the National Metrology Institutes (Batista et al., 2017) to calibrate syringe pumps. This relies on weighing the mass of water delivered by an infusion pump during a determined time. The flow rate is then determined by the quotient of the mass of reference liquid, usually water, and time interval, including some corrections following equation 1.

$$Q = \frac{1}{t_f - t_i}\left[\left((I_f - I_i) - (\delta m_{buoy})\right) \times \frac{1}{\rho_w - \rho_A} \times \left(1 - \frac{\rho_A}{\rho_B}\right) \times [1 - \gamma(T - 20)]\right] + \delta_{evap}$$

Although evaporation is kept to a minimum value, via technical means, the determined mean evaporation rate (dQevap) is required as a correction term in the volume flow rate (Q). Other contributions to the model are: the density of the reference liquid, i.e. water (rW); the time interval of the weighing, i.e. the final time (tf) minus the initial time (ti); the mass of the displaced reference liquid, i.e. the difference between the final (If) and the initial (Ii) indication of the balance; density of the air during the tests (ρA); density of the mass standards used to calibrate the balance (ρB): the water coefficient of thermal expansion(γ) and the temperature of the water during the tests (T). The term dmbuoy accounts for the buoyancy contribution of the dispensing needle immersed in the weighing vessel.

The calibration setup used during MeDD project consisted of a flow generator, connected to a device under test (DUT) upstream of a collecting vessel standing on an analytical balance (Figure 1). The balance measures the mass of the displaced liquid by the flow generator, which is precisely timed to calculate the Δm/Δt quotient.

Figure 1 – Gravimetric method setup.



During the calibration process the water to be used and the instrument to be calibrated must be at the same temperature. Therefore, the instrument should be placed in the testing room approximately 12 hours before the beginning of the calibration.

The temperature of the water used should not vary more than 2 °C during the tests.

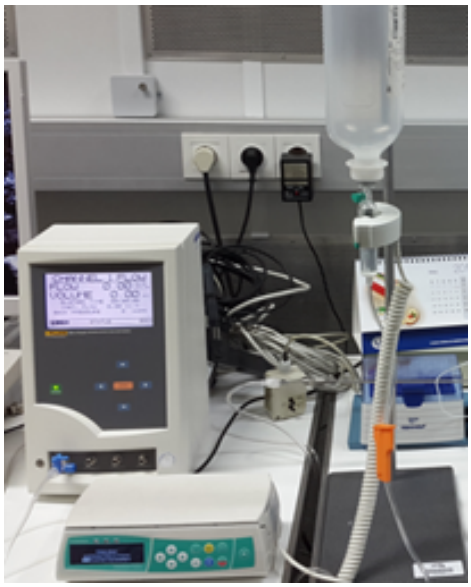
The measuring instruments (infusion pump) should be calibrated in at least three points so as to guarantee the determination of the measurement error measurement in the entire working range.

The instruments used, namely the pressure, temperature and humidity sensors, as well as the balances should have the necessary characteristics to carry out calibration and should be calibrated.

4.2. Comparison method

The other calibration method, used to determine the flow rate of an infusion pump, mainly by the hospitals maintenance offices, is a comparison method, and thereby considered as a secondary calibration method. This method consists on comparing directly the flow generated by the infusion pump under calibration with the flow recorded by an Infusion Device Analyser (IDA) (figure 2).

Figure 2 – Comparison method setup.



Before starting the calibration all the apparatus under test (syringe pump to be calibrated and the IDA) and the reference liquid should reach as close as possible to the reference temperature of 20 ºC (by 12 hours).

During the calibration, the temperature of the water and the air temperature, relative humidity and atmospheric pressure should be continually measured and/or recorded.

The syringe (normally disposable) to be used in the syringe pump is filled with ultrapure water. Before attaching the Teflon tube and mounting it on the pump, the air bubbles should be removed by inverting the syringe so that the nozzle lumen is uppermost and depress the plunger. The line should then be filled by running the syringe pump at a high rate until a steady flow of drops comes out at the end of the tube. Finally, the tube is connected with the IDA. The target flow is then programmed in the syringe pump.

Data acquisition of IDA begins after 10 minutes of steady flow and over at least 15 minutes. The data can then be directly recorded by software or read at the display as the average flow rate. The described times were obtained by experimental approaches in the Volume and Flow Laboratory of the Portuguese Institute for Quality.

The water temperature should be recorded at the beginning and at the end of the calibration, and not vary more than 2 ºC.

Concerning the measuring instruments (infusion pump), it should be calibrated in at least three points so as to guarantee the determination of the measurement error in the entire working range.

The IDA should be calibrated by accredited laboratories providing the traceability to the International System of units.

The comparative method, being a secondary method, has lower accuracy and larger uncertainty when compared with the gravimetric method, especially at flow rates lower than 10 mL/h, but it has the big advantage that can be used in clinical environment.

5. Conclusions

In order to determine the flow measurement error of an infusion pump, it is necessary to calibrate it by gravimetric or comparative method. The gravimetric method, being a primary method has a better accuracy and smaller uncertainty, but only allows the calibration of the instruments in a laboratory environment. The comparative method, of less accuracy, allows the calibration of the instruments in the hospitals, being sometimes the only possible option considering the number of infusion devices existing in each health care facilities.

The calibration of the infusion pumps ensures traceability to the International System of Units and determines their metrological conformity. This depends, therefore, on the maximum permissible error defined by the user, based on reference documentation, manufacturer's information or considering the practical use of this measuring instrument. It should also be noted that the calibration of infusion pumps must always be performed with the accessories used by the entity or service to which the instrument is affected.

References

IEC 60601-2-24:2012 *Medical electrical equipment - Part 2-24: Particular requirements for the safety of infusion pumps and controllers.*

Sectoral Comission on Health. (2018), *Metrology in Health. Good Practices Guide,* Part II-Chapter III Infusion Pumps. 1st edition. ISBN 978-972-763-172-8.

Instituto Português da Qualidade (2012), *Vocabulário Internacional de Metrologia.* 1ª Edição Luso-Brasileira. IPQ. ISBN 972763006.

Lucas, P., Klein, S.(2015), Metrology for drug delivery, *Biomedical Engineering,* **60(4)**.

Bissig, H., Petter, H.T., Lucas, P., Batista, E., et al. (2015), Primary standards for measuring flow rates from 100 nl/min to 1 ml/min – gravimetric principle, *Biomedical Engineering,* **60(4).**

ISO 7886 - 2 *Sterile hypodermic syringes for single use - Part 1: Syringes for manual use*.

Batista, E., H Bissig, H., Petter, H.T., Lucas, P et al. (2015), Assessment of drug delivery devices, *Biomedical Engineering,* **60(4).**

BIPM et al. (1995), *Guide to the Expression of Uncertainty in Measurement (GUM)*, 2nd ed., International Organization for Standardization, Genève.

Batista, E., Godinho, I., Ferreira, M., Lucas, A.. (2017), Comparison of infusion pumps calibration methods, *Measurement Science and Technology,* **28 (12).**

Authors Profiles:

Elsa Batista has a master degree in Analytical Chemistry from the Faculty of Science of Lisbon University– Portugal, Lisbon in 2007 and a degree in Applied Chemistry from the Faculty of Science and Technology, FCT, NOVA University of Lisbon in 1999. She is the head of the volume laboratory of the Portuguese Institute for Quality since 1999. She is also the contact person and Chair of the Technical committee for flow of EURAMET.

Maria do Céu Ferreira holds a Ph.D. in industrial engineering from the Faculty of Science and Technology; FCT, NOVA University of Lisbon. Since 1994, she is Metrologist at the Metrology Department of the Portuguese Institute for Quality (IPQ), having developed its activities in the areas of scientific and applied metrology, managing multi-disciplinary teams related with standards and measurements. Since 2008, Maria is in charge at Legal Metrology as Quality Manager and Assessment of the Metrological operations and verification bodies. Meanwhile, she founder and is responsible for the Portuguese Health Metrology committee, hosted by IPQ. Maria is a designated member of the Metrology Board of the Portuguese Engineers order and is Professor in the Faculty of Engineering at the Lusófona University of Lisbon. Maria is the Portuguese member of the International Board of Experts of the International Organization of Legal Metrology and is an active member of OIML, WELMEC and ISO. Her research interests include all health metrological applications for ensuring traceability and new quality assured measurement methods.

Measuring counted fractions in healthcare

L R Pendrill
leslie.pendrill@ri.se
RI.SE Metrology, Göteborg, Sweden

Jeanette.melin@ri.se
RI.SE Metrology, Göteborg, Sweden

Abstract:

In establishing metrological quality assurance in healthcare, one major hurdle is the correct treatment of ordinal data typical of questionnaires, performance tests and other categorical data collected widely in care. Despite being known well over a century, there are still many examples of measurements in healthcare – for example, (i) on-line tables of percentage performance indicators (e.g. fraction of patients seeing a doctor within seven days) and (ii) correlation plots for Alzheimer sufferers of cognitive scores against biomarker concentration – where the 'counted fraction' distortion of scales is not compensated for. The Rasch form of generalised linear model not only handles counted fraction ordinality but also enables separation of object and instrument attributes (such as task difficulty and patient ability) essential for metrological restitution in measurement systems in healthcare. A perspective is given of a new kind of certified reference material employing causal Rasch models in terms of construct specification equations for metrological item banking in the social sciences. This is part of the response to a recent call for: 'a new international body to bring together metrology, psychometrics, philosophy, and clinical management to support the global comparability and equivalence of measurement results in patient centred outcome measurement to improve healthcare'.

Keywords: Metrology; Quality assurance; Rasch Measurement Theory; Psychometrics

1. Counted fractions in healthcare

Quality assurance of products and processes of all kinds rely on the availability of quality-assured measurement. To date, many observations in healthcare, as well as in other similar areas such as sustainability, material testing, etc., are considered to lie 'off the scale' of quantitative measurement. Alongside issues concerning the adoption and implementation of regular SI quantities and units in healthcare (Ferreira & Matos, 2015), developing and implementing quality-assured measurement of categorical and performance data therefore present considerable challenges.

'Counts' (non-negative integers), of for example the number of pills, play a key role in medicine – making sure that the patient takes the correct dose, for instance. Whether 'counts' (non-negative integers), of for example the number of pills, can be considered as 'dimensionless' quantities in the international system of measurement units (SI) for the purpose of quality assurance is still an active subject of debate in the international literature (Flater, 2017). Starting from the concept of 'dimension' of a quality, Kogan (Kogan, 2014) argues that "quantity for which all exponents of factors corresponding to base quantities in its quantity dimension are zero", is preferably called a quantity of dimension one rather than dimensionless. The unit of dimension one can for example be the counted entities, such as pills.

Yet more qualitative are 'counted fractions' (bounded by zero and one), but these are common and play essential roles as performance metrics for healthcare services, ability tests, customer satisfaction, and decision risks caused by uncertainty, and are the main subject of this paper.

2. Quality assurance of counted fractions

2.1 Quality-assurance standards

Current examples of areas where metrological requirements in general are set - for traceability and measurement uncertainty - but where metrological references are yet to be established in the healthcare sector include:

- EN 15224:2012 Healthcare services - Quality management systems;
- ISO 13485:2003 Medical devices - Quality management systems -- Requirements for regulatory purposes;
- ISO 112401 Health informatics - Identification of medicinal products — Data elements and structures for unique identification and exchange of units of measurement;
- FDA 2009 Patient-reported outcome measure for verifying new / improved drugs.

Common to most quality-assurance standards following the well-known ISO-9000 series are requirements on quality assured measurement, typically in paragraph 7.6 (Figure 1, quoted here in the version found in the Healthcare service standard EN 15224:2012).

Figure 1. Paragraph 7.6 of the standard EN 15224:2012 Healthcare services – Quality management systems.

Control of monitoring and measuring devices

The organisation shall determine the monitoring and measurement to be undertaken and the monitoring and measurement devices needed to provide evidence of conformity of product (healthcare service or other healthcare product) to determined requirements…. Where necessary to ensure valid results, measurement equipment shall:

a) Be calibrated or verified at specific intervals, or prior to use, against measurement standards traceable to international or national measurement standards…

Metrology – quality-assured measurement – is often defined in terms of two key concepts:

- Traceability, which provides measurement results which are comparable over space and time. This in turn is a pre-requisite for assuring that products and processes have comparable properties, necessary for interoperability and transparency;
- Declared measurement uncertainty, which in turn allows the risks of incorrect decision of conformity of products and processes to be assessed and minimised.

A principal challenge is that there are few, if any, recognized metrological standards to establish traceability when assuring human-based quality, for instance to assure that patients can expect the same level of healthcare wherever it is provided. Healthcare has been described as a "$1 trillion per year industry without a clear measure or definition of its main product" (Heinemann et al., 2006).

2.2 Counted fractions

The challenges of handling counted fractions have been known for many years. A classic quote is: "Beware of attempts to interpret correlations between ratios whose numerators and denominators contain common parts" (Pearson, 1897).

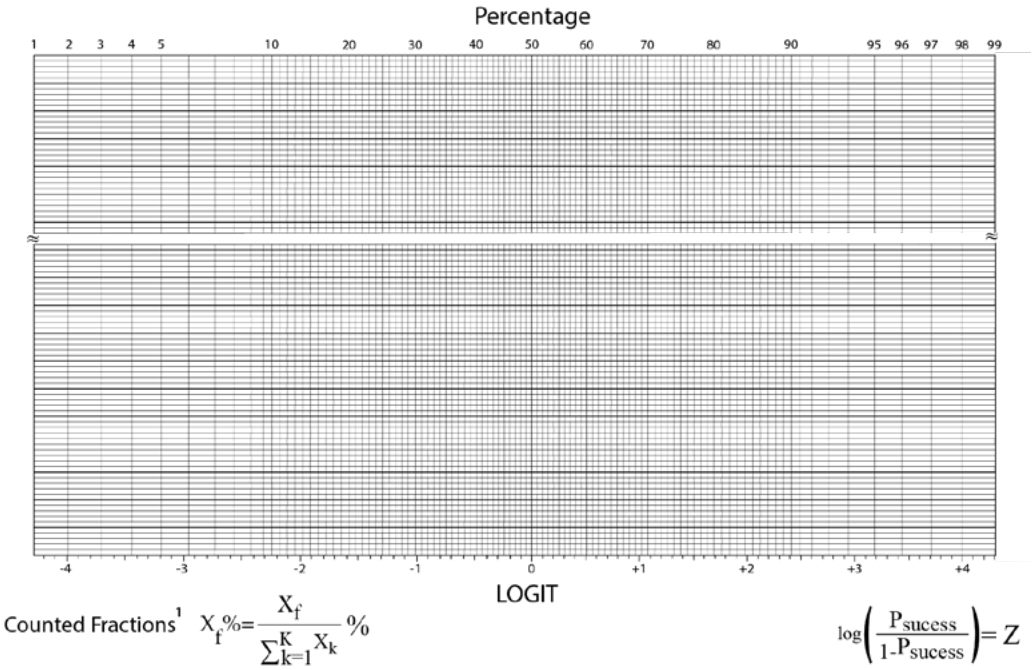
The kind of phenomenon characteristic of counted fractions being referred to has been described by Tukey (Jones, 1984):

> First, our experience-molded intuitions tell us clearly that it is not a mode where equal numerical changes correspond to equally important changes. A change of 5 % is not equally important across the scale. The difference, for almost all purposes except voting (where it is routine to use the qualifying expression "percentage points"), between 1 % and 6 % is very much more important than the difference between 48 % and 53 %. Once we break down our idea that "percentages are the proper mode," we come to feel quite clearly that we need to open out the scale for extreme percentages, as compared with percentages near 50 %.
>
> As a general consequence we should expect that scales which have a finite range are likely to give us trouble, unless our observations tend to be safely away from any ends which are present. Hence the fact that percentages go only from one end (at 0 %) to another (at 100 %) suggest that, whenever even moderately extreme percentages are likely to occur, we are likely to "stretch the tails", while, if really extreme percentages occur, we may have to stretch hard enough so that there are no ends (at any finite values) (Tukey quoted by Jones 1984).

Mathematically, in the counted fraction expression, $X_j\%=X_j/(\sum_i X_i)$ eq. (1), the presence of the amount X of component j appearing in both the numerator and denominator means – and increasingly where is either large or small compared with the other components – that any error in will be correlated with the other components, since of course there is the boundary condition $\sum_j Xj\%=100\%$. It is straightforward (Pendrill, 2018), using the method of Lagrange multipliers with this boundary condition as constraint, to derive the so-called 'link function' between ordinal, counted fraction raw data, such as percentages, p, and a more quantitative, linear scale: $l=\ln(p/(100-p))$.

In the days before computers in the first half of the 20th Century, such non-linearities at the scale extremities – both the high and low ends – although known to be linearisable through so-called logistic ruling, converting percentages p into logits, l, were arduous to calculate, and it was common to use monographs, such as exemplified in Figure 2.

Figure 2-Logistic ruling (reproduced and adapted from Tukey (Jones, 1984).



These days, Generalised Linear Models (GLM) which can handle ordinal scale properties can be readily implemented. Well-known examples are the Rasch psychometric programs which can be used to treat measurement system response, Psuccess, where mathematical distances between categories are not known exactly. Being based on the principle of least action and the Lagrange multipliers method, GLMs are not unique to human-based perception, but indeed apply to a wide range of ordinal and nominal data. Note however that counted fractions lead to an ordinal scale, but not all ordinal scales are related to counted fractions.

As will be seen in the case studies below, examples may be readily found in modern healthcare where data analysis is not done correctly. The consequences of incorrect decisions about care arising from improper treatment of counted fractions have in many cases not been analysed.

3. Restitution of counted fractions and metrological references

The counted fraction issue is an important part of the on-going discussion about how many concepts of traditional engineering and physical metrology can be extended to apply also to measurement in the social sciences (Cano et al., 2016, 2017).

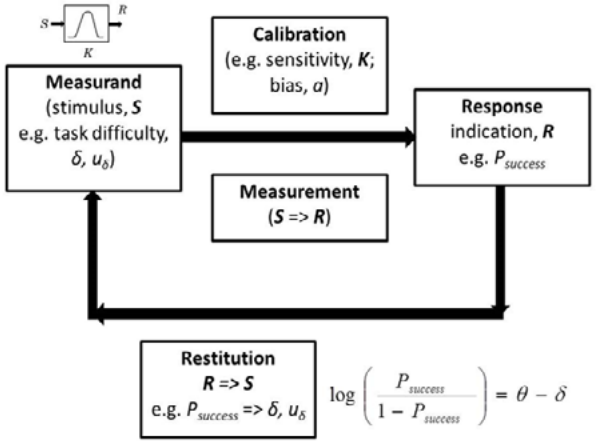
When attempting a reconciliation between metrology in the physical sciences and measurement practices in the social sciences, a good start is to regard a human being (or other 'probe', when perceiving any 'entity' in the broadest sense) as a Measurement Instrument, as suggested in the pioneering multidisciplinary European Project 'Measurement the Impossible Network' (MINET; Pendrill et al., 2010). 'Perceptive' measurement system response is however not a simple instrument indication, but instead extends to the decision risks caused by uncertainty. Final restitution of the measurand (e.g., task difficulty) from this performance metric response in a form suitable for metrological quality assurance requires specific treatment.

Apart from handling counted-fraction ordinality, Rasch measurement theory (i.e., postulating the link function $z = l = \theta - \delta$) is unique amongst GLM in additionally enabling a separation of a 'probe' attribute θ and a 'target' attribute δ. This attribute separability is essential to underpinning measurement traceability and uncertainty, in the same way as the sensitivity of a weighing scale needs to be calibrated separately from the stimulus value of a mass (Pendrill, 2018).

In a conventional measurement system, where the input signal is measurement information from the measurement object, then the sought-after value of the quality characteristic of the object can be restituted from the observed output. The restitution process consists of inverting the observation equation of the measurement system to estimate the stimulus S in terms of the system output R, and other terms. An assumption is of course

that system factors, such as the sensitivity K, remain unchanged from when calibration of the measurement system was performed. A simple example is a measurement system where the instrument sensitivity and there is an offset (bias), a, in the output, R. The formula for restitution of an unknown input, S, that is the stimulus value of the measurement object in this case is: , which might need to be evaluated at individual input levels if either sensitivity and/or bias vary with level.

Figure 3- Observation and restitution for performance metrics.



Source: Pendrill (2018)

The analogous separation of attributes of the measured object ("item") from those of the instrument ("person") measuring them, achieved with the Rasch model, brings invariant measurement theory to psychometrics.

A full picture of the measurement process when Man acts as a measurement instrument [Figure III], presents the process, step by step, from the observed indication (a performance metric, e.g. probability of success, Psuccess, of achieving a task, and restitution with Rasch Measurement Theory, through to the measurand (e.g. task difficulty) in a form suitable for metrological quality assurance:

In addition to the regular elements of any measurement system, such as sensing, signal conversion and processing, and data processing (Bentley, 2005), an important fifth category has to be added; namely, a decision-making element which models the performance metrics typical of measurements in the social sciences. Most measurements are not made solely for the sake of measurement, but because decisions are to be made about something – a product, a process, service or phenomenon - based on the measurements: Decision-making: algorithm producing an output on a categorical scale: the result of a decision, such as the binary, dichotomous response if the measured response Tm is above or below a specification limit TSL:

$$R = \begin{matrix} 0 \\ 1 \end{matrix} if \begin{matrix} T_m \leq T_{SL} \\ T_m > T_{SL} \end{matrix}$$

For such categorical response cases, including the important decision-making response measurement system 'accuracy' – as a performance metric - will be identified with decision-making ability: Accuracy (decision-making) = response categorisation – input (true) categorisation, where is a metric of measurement system performance in terms of the probability of making the 'correct' decision:

$$z = \theta - \delta = \log\left(\frac{P_{success}}{1-P_{success}}\right)$$

4. Case studies

Tukey (Jones, 1984) gives a number of examples where these effects of counted fractions can occur:

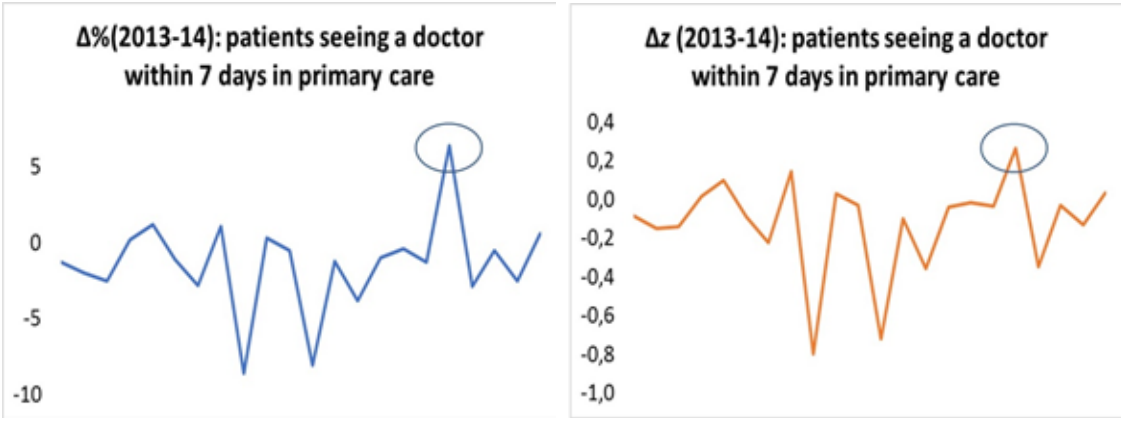
> «Counting sheep and goats, and reporting on the relative number of goats, still typifies much of behavioural science … As a consequence, the behavioural sciences have a very strong continuing interest in modes of expression of counted fractions, although they may appear to be unaware of this interest …. Experience with a rather wide variety of relative-number problems, varying from "how many were affected at this dose" to "how many of the pebbles are quartz", indicates that further analysis proceeds smoothly and thoroughly when other modes of expression are used instead of "percentages".»

Despite knowledge and available tools, counted fractions are still commonly ignored. As Aitchison (1982) notes in the context of the related field of compositional data analysis, so-called "describers" merely state: 'After all we are simply describing the data set in summary form, not analyzing it.' Another group are the "wishful thinkers": 'No problem exists or, at worst, it is some esoteric mathematical statistical curiosity which has not worried our predecessors and so should not worry us. Let us continue to calculate and interpret correlations of raw components. After all if we omit one of the parts, the constant-sum constraint no longer applies. Someday, somehow, what we are doing will be shown by someone to have been correct all the time.' (Aitchison, 1982).

4.1 Performance metrics in healthcare services

From only a few years ago, where annual pdfs were published on-line with detailed statistics of performance metrics for various healthcare services, these days websites are appearing where colourful tables warn in red or highlight in green negative and positive changes in regional healthcare services, almost in real time. But some of these have to be taken with a 'pinch of salt'.

•Figure 4a,b. Changes between 2013 and 2014 in the fractional number (y-axis) of patients seeing a doctor within 7 days for a range (x-axis) of regional primary care providers (a) in % terms according to eq. (1); (b) in logits, z, according to eq. (2) (SALAR, 2018).



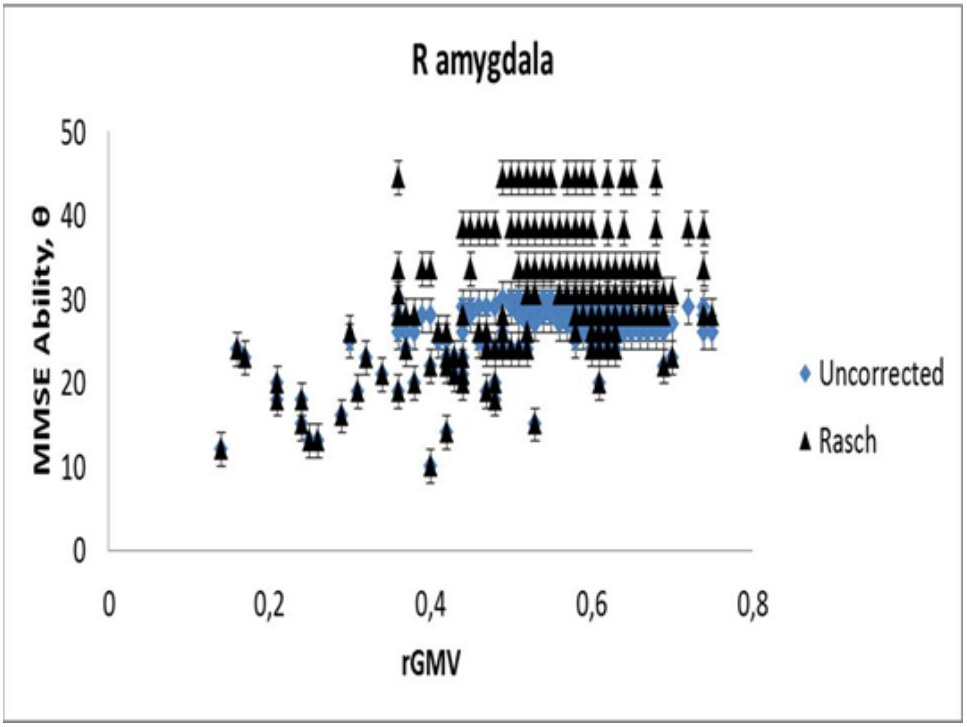
Source: SALAR (2018)

Figure IVa apparently shows a dramatic improvement by + 6 % for one particular regional primary care provider between the years 2013 and 2014 in the fractions of patients seeing a doctor within 7 days when the data is presented in percentages, according to eq. (1) (SALAR, 2018). But it is easy to reveal that this is just a counted fraction mirage, and is not matched by a corresponding change in logit metric z, according to eq. (2) (Figure IVb). The explanation is that one has to account for two separate factors; not only the change in percentage points but also the absolute level in percent for each quantity compared. In this example, the change for the regional provider in question spanned a range of absolute values 85 % - 91 % between the years 2013 and 2014, where each counted fraction scale has appreciable and different non-linearity according to eq. (1) at each absolute level. The corresponding logit change [Figure IVb] shows that this specific region's apparent improvement in care service provision performance is not significantly better than other regions, and indeed lies within intra-regional scatter.

4.2 Cognitive studies of Alzheimer's disease patient

As a second example where counted fractions have to be corrected for is regarding person-centred outcome measures, such as evaluations of abilities or leniency. Recent studies of possible correlations between neurodegeneration in patients suffering from Alzheimer's disease – specifically failing cognitive ability – and brain atrophy are part of the on-going EMPIR HLT04 NeuroMet project. A recent application of the Mini Mental State Examination (MMSE) test can be found of potential correlations between cognitive ability and brain atrophy studied by Dinomais et al., (2016) - their original ("uncorrected") data is plotted in Figure V. MMSE scale distortions arising from the counted fraction effect, surprisingly, do not seem to have been accounted in an otherwise extensive literature (Klein-Koerkamp et al., 2014) covering advanced correlation studies between cognitive and biomarkers of neurodegenerative diseases. But these can be corrected for in a metrological manner thanks to the Rasch invariant measure theory, which for MMSE has already been investigated by Hughes et al. (2003).

Figure 5- Correlation plots of cognitive ability (MMSE, dependent variable) versus regional grey matter volume (rGMV, independent variable).



original data (Dinomais et al., 2016); X Rasch corrected for distortion shown in Figure 4 (Hughes et al., 2003), and this work. Standard uncertainties are indicated for MMSE ability scores (Pendrill, 2018).

5. Metrological item banks

The need for better coherence between to date parallel initiatives in the metrology of healthcare from the traditional metrological organisations (predominantly in the physical and chemical sciences) and healthcare researchers has been highlighted recently. Cano et al. (2017) for instance: '….propose a new international body to bring together metrology, psychometrics, philosophy, and clinical management to support the global comparability and equivalence of measurement results in patient centred outcome measurement to improve healthcare'. In laboratory medicine, the standardization of measurements is of course a high priority, with the goal of comparability of results obtained using routine procedures (Jones & Jackson, 2016; Hallworth et al., 2015). 'A determination to accept patient outcome and patient experience as the primary measure of laboratory effectiveness' is a recommendation of a recent International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) Task Force on the Impact of Laboratory Medicine on Clinical Management and Outcomes. The IFCC in turn belongs to the Joint Committee for Traceability in Laboratory Medicine as a link to the metrology community.

The present work indicates that certified reference 'materials' (CRM) can apparently be synthesised in terms of construct specification equations, i.e., experimental intervention/manipulation on either person attribute (e.g., ability) or object attribute (e.g., difficulty) or both simultaneously, that yield successful prediction of the observed outcome (count correct). These causal Rasch models relate task difficulty and patient ability to explanatory variables such as the test sequence entropy and brain atrophy. This means more than merely running data through Rasch calibration software. The ability of the Rasch approach to yield separate and objective measures of task difficulty and person ability, for instance, is essential in establishing sets ("item banks") of metrological references based on proper use of the psychometric Rasch model (Choppin, 1968; Pesudovs, 2010). The bottom line is that 'counted fractions' do indeed belong to an extended quantity calculus, thus enabling quality-assurance of many essential measurements in healthcare.

Acknowledgments

This work has been performed as part of the 15HLT04 NeuroMet project which belongs to the European Metrology Programme for Innovation & Research (Horizon, 2020), jointly funded by the EMRP participating countries within EURAMET (www.euramet.org) and the European Union.

Additional funding has been provided by VINNOVA, the Swedish Innovation Agency, in the project: Quality-assured measurement in Swedish healthcare, Reg. no. 2017-05011. The aim of that project is: "Care is increasingly delivered with the entire health condition of the person in focus, rather than just the disease. In this context, there is a growing need for quality-assured measurement of categorical measurements, but typical human responses, such as opinions on ordinal measurement scales, are more difficult to handle quantitatively. The purpose of the present project is therefore to start work on establishing a national metrology institute (NMI) for quality assurance of categorical measurements in healthcare. In the long run, the goal is to become a leading NMI in the area."

Thanks are also due, particularly, to:

- Members of the EMPIR NeuroMet consortium;
- Stefan Cano, Modus Outcomes, Boston (USA) & Stotford UK;
- William P. Fisher, Jr., Ph.D., Research Associate, BEAR Center, Graduate School of Education, University of California, Berkeley, CA (USA) & Principal, LivingCapitalMetrics Consulting.

References

Aitchison, J (1982), The statistical analysis of compositional data, *J. R. Statistic. Soc.* **44,** 139 -77.

Bentley, J P (2005), *Principles of Measurement Systems.* 4th ed. London: Pearson Education Limited.

Cano, S J, T Vosk, L R Pendrill, A J Stenner (2016), On Trial: the Compatibility of Measurement in the Physical and Social Sciences, *Journal of Physics: Conference Series* 772 012025, IMEKO2016 TC1-TC7-TC13, doi:10.1088/1742-6596/772/1/012025.

Cano, S J, L R Pendrill, S P Barbic and W P Fisher, Jr. (2017), Patient-Centred Outcome Metrology for Healthcare Decision-Making, *Joint IMEKO TC1-TC7-TC13 Symposium: Measurement Science Challenges in Natural and Social Sciences,* Rio de Janeiro, Brazil, July 31st - August 3rd 2017.

Choppin, B (1968), An item bank using sample-free calibration, *Nature* **219,** 870 -2.

Dinomais, M, S Celle, G T Duval, F Roche, S Henni, R Bartha, O Beauchet and C Anweiler (2016), Anatomic Correlation of the Mini-Mental State Examination: A Voxel-Based Morphometric Study in Older Adults, *PLOS ONE* | DOI:10.1371/journal.pone.016288.

EN 15224:2012, *Health care services – Quality management systems – Requirements based on EN ISO9001:2008,* European standard.

FDA (2009), *Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims,* U.S. Department of Health and Human Services, Food and Drug Administration, http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm.

Ferreira, M and A Matos (2015), Metrology in health: a pilot study, *Journal of Physics: Conference Series* **588,** 012018.

Flater, D (2017), "Redressing grievances with the treatment of dimensionless quantities in SI", *Measurement* **109,** 105–11.

Hallworth, M et al., (2015) Current evidence and future perspectives on the effective practice of patient-centered laboratory. *Med. Clin. Chem.* **61** 589–99.

Heinemann, A., W. Fisher, Jr., and R. Gershon (2006), Improving Health Care Quality With Outcomes Management, *J. Prosthet. Orthot.,* **18,** no. 1S, pp. 46-50.

Hughes, L F, K Perkins, B D Wright and H Westrick 2003 Using a Rasch scale to characterize the clinical features of patients with a clinical diagnosis of uncertain, probable, or possible Alzheimer disease at intake *J. Alzheimer's Dis.* **5** 367–73.

ISO 13485:(2003), *Medical devices - Quality management systems -- Requirements for regulatory purposes,* International standard.

Jones G, and C Jackson 2016 The Joint Committee for Traceability in Laboratory Medicine (JCTLM)—its history and operation, *Clin. Chim. Acta* **453** 86–94.

Klein-Koerkamp, Y, R A Heckemann, K Ramdeen, O Moreaud, S Keignart, et al., (2014), Amygdalar Atrophy in Early Alzheimer's Disease, *Current Alzheimer Research* **11,** pp.239-252, https://hal.archives-ouvertes.fr/hal-00959516.

Kogan, J. (2014), An Alternative Path to a New SI, Part 1: On Quantities with Dimension One, https://web.archive.org/web/20160912224601/http://metrologybytes.net/PapersUnpub/Kogan_2014.pdf.

Pendrill, L R, B Berglund et al., (2010) Measurement with Persons: A European Network, NCSLi Measure, (June 2010) 5, 42 – 55.

Pendrill, L R (2018), Assuring measurement quality in person-centred healthcare, *Measurement Science & Technology,* Volume 29, Number 3 , 034003 special issue Metrologie 2017, https://doi.org/10.1088/1361-6501/aa9cd2.

Pesudovs, K (2010),Item banking: a generational change in patient-reported outcome measurement, *Optom Vis Sci.,* 87(4) pp 285-93. doi: 10.1097/OPX.0b013e3181d408d7.

Tukey, J W [Chapter 8, Data analysis and behavioural science; quoted in "The collected works of John W Tukey, Volume III, *Philosophy and principles of data analysis: 1949 – 1964*", ed. L V Jones, Univ. North Carolina, Chapel Hill, Wadsworth Statistics and Probability Series. Wadsworth & Brooks/Cole, Belmont (1985).

SALAR (2018), Vård i siffror, (Swedish: Care in numbers), https://vardenisiffror.se/registry/registry?descriptionSource=V%C3%A4ntetider%20i%20v%C3%A5rden,%20Sveriges%20Kommuner%20och%20Landsting .

Authors Profiles:

Prof. L R Pendrill, Ph D Atomic Physics 1978, has until recently been R&D Director at Measurement Technology Division and the National Metrology Institute of SP at Borås. He plays a leading role, nationally and internationally, recently as Chairman (2009-12) for EURAMET (www.euramet.org), the European association of national metrology institutes in Europe. Pendrill is currently Chair of ISO/TC12 Quantities & Units. He develops new quality-assured measurement methods in which Man is focus by measuring PCOM (e.g. perceived comfort), as well as optimised measurement uncertainty, where the effects of measurement are weighed against the effects of product evaluations.

Dr. J Melin has the last years been working with measurements in health care. She is now working at the division for Metrology at RISE Research Institute of Sweden, where she is included in several project where person-centered metrology is used and further developed. During her PhD project she took part in the development of the PPRQ at the Gothenburg Centre for Person-Centred Care (GPCC). After her dissertation in December 2016 she has both continue the development of PPRQ and her skills in metrology.

Metrology in Medical Field

Baki Karaböce
baki.karaboce@tubitak.gov.tr, bkaraboce@outlook.com
TÜBİTAK UME, National Metrology Institute

Abstract:

Medical device calibration and test are one of the emerging, important and critical issues in the field of metrology. The traceability in medical devices and measurements are not powerful enough as much as the traceability in technical and military calibration and measurements. Patient safety is a must for the medical device industry and applications in health sector. Therefore all measurement devices used in medical field must be controlled periodically and all measurements must be standardized as a quality control regimen that guarantees the reliability of medical devices. Test, measurement and calibration of bio-medical equipment are becoming increasingly significant for manufactures, when accuracy in diagnosis and effectiveness in treatment are required as well as patient safety.

In this paper, medical metrology concept has been introduced and critical cases have been outlined. Current situation has been summarized in all over the world by the light of metrology institutes and international organizations. Measurement parameters in medical devices have been investigated and calibration systems, traceability and standards have been given. It is expected that this manuscript will give a clear vison for the importance of accurate measurements in medical field and guide for medical metrology studies.

Keywords: Measurements, medical device, medical metrology, traceability, uncertainty

1. Introduction

Medical metrology can be described as measurements involving measurement, control, verification and calibration activities of measurement, analysis, diagnosis, imaging and treatment devices. Devices used in this field have critical importance because they are used on human and they are multifunctional.

Medical decisions about diseases are usually based on the clinical findings including medical examination and the results of statistical studies that have been obtained from the patients over the years. Medical measurements are important for creating of clinical findings and generating consistent statistical data from large number of patients (Report, 2013; Schreyögg, 2009; Zimmerman, 2014).

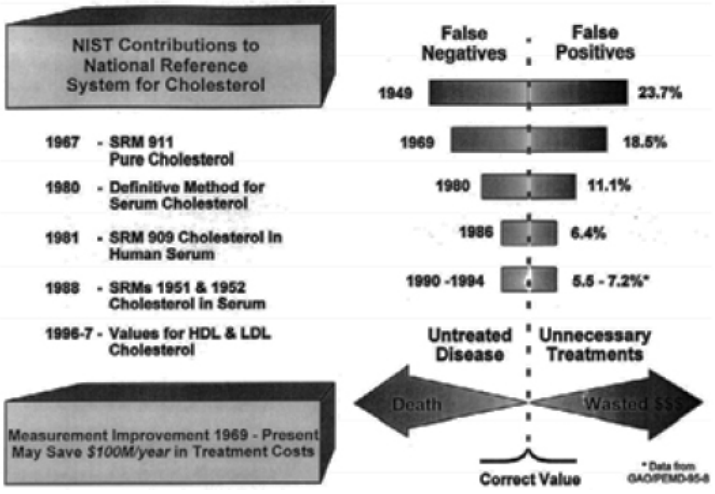
Medical measurements are actually based on the research studies conducted by the work of national metrology institutes. A reliable measurement system can be created by combining metrological instruments and quality assurance systems. Calibrations of the devices and systems used for transferring the measurement values, using of scientifically accepted measurement and calibration methods in measurements and control procedures required by legal metrology are metrological tools that must be used in the system. Objective and transparent system based on quality criteria can be established by using the internationally recognized standards of quality assurance (Karaböce, 2016).

In the field of health, the diversity of devices used in imaging, diagnosis and treatment has been increasing in recent years and its usage is becoming widespread (Squra, 2015; Tooker, 2005; Zaitseva, 2017).

These systems range from a simple thermometer to computerized imaging systems, clinical measurement devices or computer-controlled, highly precise surgical robots. According to the findings and examination findings obtained with these systems, the treatment steps and methods are determined by the physician. Data obtained from measurement systems in audits may be more important than findings from examinations. For example, a hernia that cannot be identified in the examination findings of medical doctor can appear in MR (magnetic resonance) image.

As a result of the measurements, analyzes and tests made in the field of clinical metrology, the direct treatment method can be determined. For example, if a diabetic patient has a high blood sugar level, it is highly likely that he or she will begin insulin therapy. Failure to measure glucose in the blood correctly will result in unnecessary treatment and/or delay in necessary treatment. Damage risk to life on the wrong measurement of glucose in the blood cannot be measured in addition to high amount of expenses. More accurate or small uncertainty in cholesterol in the blood measurements will guarantee the less expenses and correct treatment as can be seen in Figure 1 (NIST; Semerjian, 1999). Definitive and standard methods and development of serum cholesterol standard reference materials (SRMs) will result in saving of treatment costs for misdiagnosed patients and additionally lives through timely and accurate diagnosis.

Figure 1 - Development of accuracy in glucose in the blood measurements.



Source: Semerjian, (1999)

Regular maintenance/repair as well as testing, measurement, verification and calibration must be carried out at regular intervals to ensure correct operation of medical devices during use. Some of the devices and apparatus used for testing, measuring, verifying and calibrating are commercially available. However, it is very important that institutions and organizations that provide services in the field of health as well as all areas should be able to monitor the devices used for diagnosis and treatment to national and/or international standards and to establish a measurement association within the country. It is necessary to carry out studies for the establishment and research of the necessary infrastructure for this traceability (Baura, 2012; Shirmohammadi, 2016).

In order to emphasize the importance of instrument calibrations used in the medical field and to ensure correct calibration, training programs are provided in some universities, public and private institutions, in limited numbers and narrow scope. These training programs should be disseminated. There should be absolute medical metrology among the training programs and uncertainty calculations in the measurements. Training programs should be given to all interested persons, including hospital administrators, calibration laboratory staff and device users.

A regulation on testing, control and calibration of medical devices will create an appropriate environment in order to establish traceability of standards and reliability in medical measurements. In particular, audits should cover technical areas. For example, aspects such as calibration, compliance with a specific standard, traceability of the devices used should be controlled. It is foreseen that the accreditation institution and the pharmaceuticals and medical devices agency should act in coordination with the accreditation of the laboratories that perform medical device calibration.

Collaborative platforms for the collaboration of public institutions, universities and industrial organizations for the design and production of devices such as calibrators, phantoms, analyzers in the medical field are more widely and effectively used.

Scheduled works in the field of medical metrology by national metrology institute (NMI) or designated institute (DI) can be listed as follows:

- Making preliminary study to ensure traceability in national and international standards for diagnostic and therapeutic devices which are serving in the field of health;
- Preliminary studies for the establishment of a measurement unity within the country for devices used in the health field;
- In the field of clinical metrology, investigation of required laboratory infrastructure for validation of measurement method used to verifying measurements, analyses and tests;
- In clinical metrology, planning for the production of reference materials used in chemical measurements;
- Creation of metrology awareness for universities, calibration laboratories and societies which are interested with this subject;
- Comparison and competence testing for the purpose of ensuring measurement unity between laboratories in the country.

The metrological reference values (produced for country and by joining all necessary all international comparisons required for NMI's reference standards and reference materials used in the traceability of medical and clinical devices) are aimed to adapt to the international system.

Examination of written standards, medical and clinical devices, calibration/measurement systems within the feasibility project and information obtained from interviews with related institutions, universities, hospitals and private laboratories must be compiled. With this information, works to be done, calibration/measurement systems to be installed, plans of these calibrators, phantoms and reference materials which are considered of their design and production have to be given.

2. Current Situation in Medical Metrology

Official institutions, universities, biomedical calibration laboratories and private companies operate in specific areas of medical metrology, have variable approaches in this field (Bošnjaković, 2017; Ferreira, 2011; Feldman, 2007). Some private hospitals declaring that they are testing, checking, verifying and calibrating their own devices. Medical calibration units in the Ministry of Health declare that they can test, control, verify and calibrate the medical devices in some hospitals within the capacities of their own portable system. But the traceability and reliability of medical devices is not provided by metrologically. calibrations of instruments used in calibration (such as calibrator, simulator, analyzer and phantom) must be performed by authorized (accredited) and trained persons, using reference measurement standards, according to written standards or methods defined in the literature, if possible in controlled environments.

Some good examples can be given: university hospital's devices are calibrated with the systems in the relevant laboratories where they calibrate their own devices in their systems. But some of them send devices such as simulators, analyzers, etc. used in calibration to an organization that does not have an accreditation certificate (Monteiro, 2017).

As can be seen from the above information, they are doing their own calibrations but the calibration must be done by independent third parties. When the situation is examined in terms of metrological traceability and reliability, it can be said that the calibration of the reference standards should be done by the accredited laboratories or the national metrology organization in the country.

Most hospital's devices cannot be calibrated in real sense. For example, it can be seen that the calibration requirement is fulfilled by attaching only labels and performing non-accredited measurements (Ventola, 2008; Kiekens, 2010). The reasons can be listed as follows:

- Fulfillment of calibration makes possible to receive money from circulating capital fund of the hospital;
- The ignorance or carelessness of the medical doctors or the person concerned who request for calibration;
- There is no regular audit by authorities.

Calibrations of calibrators, simulators, analyzers and phantoms which are used in the test, control, verification and calibration of medical devices are made by the distributors of those devices under the name of maintenance/repair/calibration in health institutes. This process is not correct metrologically and not within the scope of authority and accreditation. NMI must calibrate the calibrator, simulator, analyzer and phantoms with creating references to medical metrology:

- The situation has been established and it has been emphasized that the major of devices used in the medical field are not calibrated or verified metrologically;
- Considering the frequency of use, measurement accuracy and health risk, it has been agreed to assign a calibration period for each device type. In general case, hospital's surgeons are not well aware of the importance of test, control, verification and calibration of medical devices;
- Many manufacturers and distributors have no information about device calibration;
- The biomedical institute or departments of some universities are only more focused on consultancy and education but not calibration and measurement.

For this purpose, training programs should be organized by the NMI and other organizations / universities and participation to these trainings should be encouraged. Most of the calibration companies do not have sufficient coverage in the accreditation certificate.

It has been stated that in the initial installation and licensing stages of ionizing radiation devices which are used in the health field, authorized agencies for ionizing radiation are taking part but later they don't do calibration and measurements. Safety of usage of medical device must be settled in order to avoid tomography, mammography, and X-ray devices produce higher doses from expected doses. Application of excess doses must be prevented in order to break kidney stones with few sessions in lithotripsy laboratories (Hermanek, 2017; Ainsley, 2014; Villarraga-Gomez, 2017).

NMI must ensure the testing, control, verification and calibrations of medical devices in health institutes with reference standards. The suggestions made during the interviews in the field of medical metrology are listed below. The following suggestions should be made to start working at NMI in the light of the recommendations. Calibrators, phantoms and reference materials should be designed and manufactured in NMI. Trainings and workshops emphasizing the importance of medical device calibration should be organized.

Bureau International des Poids et Mesures-the Joint Committee For Traceability in Laboratory Medicine (BIPM – JCTLM) was established in 2002 for the implementation of the European Union Directive 98/79/EC with the cooperation between the Comité International des Poids et Mesures (CIPM), the International Federation of Clinical Chemistry and Medical Laboratories (IFCC) and the International Laboratory Accreditation Association (ILAC). JCTLM contributes to provision of the traceability of the measurements made in the clinical laboratories. The calibrators used by laboratories which are operating in the field of in vitro diagnostics are also inspected. JCTLM activities cover production of primary reference materials and reference measurement procedures/methods, and service for laboratory reference measurement worldwide.

The National Metrology Institute of Turkey (TÜBİTAK UME) has established the traceability of measurement quantities relevant to medicine, to integrate the measurement quantities into the international metrology system through international comparisons and to ensure measurement unity by disseminating traceability to lower level laboratories within the country or abroad through calibration, measurement and test services. The other objective of the laboratory activities is to produce research projects in line with national priorities and the stakeholder demands.

As well as ensuring measurement traceability of the devices used in medicine and reliability studies, the Medical Metrology Laboratory has implemented projects such as calibrator design and production for medical devices and certified reference materials production for autoanalyzers (TUBİTAK UME). In addition to this, the laboratory has also organized trainings for professionals that perform medical device calibrations.

Study areas of the TÜBİTAK UME Medical Metrology Laboratory include:

- •Ensuring measurement traceability of the devices used in health field;
- •Certified reference materials production for clinical measurements;
- •Calibrator design and production for medical devices;
- •Practical trainings for professionals perform in medical device calibrations;
- •Establishing systems and implementing projects for the application of ultrasonic techniques in the health field;
- •Conducting performance tests of hearing aids and headsets;
- •Calibration and Measurement Services include;
- •Patient Simulator Calibrations;
- •Defibrillator/Pacer Analyzer Calibrations;
- •Electrical Safety Analyzer Calibrations;
- •Pulse-Oximetry Analyzer Calibrations;
- •Infusion Pump Analyzer Calibrations;
- •Gas Flow Analyzer Calibrations;
- •Electrosurgey Analyzer Calibrations;
- •Hearing Aid Performance Tests;
- •Measurement Systems for Ultrasonic Applications in the Health Field;
- •Standard Reference Materials (SRM);

United States National Institute of Standards and Technology (NIST), has medical metrology activities as shown below:

•Optical imaging techniques: With this program, the development of optical imaging techniques in surgeries, in hospitals and in medical applications is planned;

•Electronic-based biosensing, single-molecule metrology;

•Engineering studies that aim to reproduce the organ structures in three dimensions;

•NIST has developed a reference phantom for using in calibration of magnetic resonance imaging systems. 100 contrast agents and reference material placed to inside to sphere as big as human skull (NIST; Chiao, 2008).

The French National Meteorological Institute-Laboratoire National de Métrologie Et D'essais (LNE), expanded his work in health metrology, certification activities are spread to 52 country, 4000 laboratories in the country reported that they provide reference measurement data on biomarker dosage every day according to the 2010's year report (LNE):

•LNE established test laboratory in 1979 year for meet the certification requirement of hospitals and medical device manufacturers in the field of health;

•LNE performs measurements on hearing aids, blood pressure monitors, eye pressure monitors and spirometers in the medical field;

•In addition, LNE's laboratories have the ability to perform the calibrations and tests on medical devices and equipment;

•LNE certifies the medical devices in accordance with ISO 13485, ISO 15378, ISO 15189 and EN 15593 standards;

•Research activities are the development of measurement systems that can detect biomarkers, the development of test methods that can provide monitoring and control of electronic blood pressure gauge, and nano virology.

German Physcalische-Technische Bundesanstalt (PTB) conducts activities in the field of medical device. Main purpose is to improve accuracy and reliability in diagnosis and treatment by creating measurement methods and test procedures and scientifically, industrially and aimed protecting the consumer. The department engages in method development for magnetic resonance imaging (MRI) and spectroscopy (MRS) as seen in Figure 2. In the Medical Physics Department new measurement and testing techniques are being developed and research is being conducted about reference material production. PTB has improved the secondary personal dose Hp (10) standard for X-ray use. The standard is designed as the volume of ionization in a 30 cmx30 cmx15 cm polymethyl methacrylate (PTB; Ittermann, 2015).

Figure 2 - MRI imaging studies at PTB

National Physical Laboratory of United Kingdom (NPL), has research and development studies in the medical field (NPL). Working areas and projects:

•Optical diagnostic tools (Optical Coherence Tomography) and optical tissue phantoms;

•Centre for Biomolecular Metrology (polypeptide structure and function);

•Calibration of ultrasound power for medical devices.

NPL has published studies of patient's treatment, which is at home for long-life and monitoring by tele-care monitoring, and studies of to ensure the reliability of devices used in therapeutic measurements in annual reports.

The Japanese national metrology institute (NMIJ) is working on standardizing medical metrology measurements. For this purpose, NMIJ has created an organization, named "Standardization of Laboratory Medicine Club" and has been carrying out studies especially in the Department of Biomedical Standards and Organic Analytical Chemistry Department (NMIJ; Nakamura, 2010; Tanaka, 2014). NMIJ carried out the studies in the field of medical metrology are medical imaging systems and production of reference materials for clinical laboratories. By organizing workshops in the field of medical metrology; NMIJ ensures that international trends in the sector and new international standards in the medical field are followed.

The South Korean Metrology Institute (KRISS) has established a center in the field of medical measurements (Center for Medical Measurements). Their primary research works are the development of national calibration systems for blood pressure gauges and thermometers which are commonly used in hospitals. Also, the required metrological studies are being carried out in order to establish the traceability of the signals which produced by ECG devices (KRISS).

2.2. International Organizations

The Joint Commission Accreditation for Health Organization (JCAHO) provides accreditation organization in "Hospital management and care services and professional management functions". This movement, which started in health care services in the United States, has been a study that attracts the attention of world countries since 1990.

Joint Commission International (JCI) that is a subdivision of JCAHO works as an institution established to accredit hospitals in the international arena. JCI's mission is to make better and improve the quality of health services in the international arena by providing worldwide accreditation services. "Accreditation Standards for Hospitals" book, which is prepared for the hospitals to be prepared for the accreditation audit by the JCI organization, includes the following headings about accreditation standards for administrative areas:

- Quality improvement and patient safety;
- Management, leadership and orientation;
- Facility management and safety;
- The quality and training of employees;
- Prevention and control of infections;
- Information management.

Approximately 190 hospitals in countries outside the US have received JCI accreditation. JCI generally conducts technical surveys at the hospital to ensure the quality and control of all operations.

Emergency Care Research Institute ECRI activities are for development of methods and standards that will remove risks from patient care and patient safety. ECRI study and research areas are about measurement/calibration systems to be established in the field of medical metrology, calibration/measurement methods and reference materials which will be developed, and the medical devices which are used in hospitals. Institute provides the accurate measurements in the desired level of uncertainty, the sustainability of these measurements through the accreditation system. As a result, ECRI accreditation system will ensure that the clinical, technical and administrative services provided by the health services are guaranteed to give more accurate and safe services.

World Health Organization (WHO), conducts international studies on community health. The United Nations Conference, gathered in San Francisco, USA in 1945, acknowledged that the health of all people has fundamental importance of ensuring peace and security in the world. And it has been accepted that organize a meeting for establishing an "International Health Organization" by Chinese and Brazilian delegates unanimity (WHO).

Some of the tasks, that organization fulfilled to achieve its organizational goals, are listed below:

- •Promote and guide research in the field of health;
- •Facilitate the improvement of the norms of teaching and training of medical staff;
- •Standardize diagnostic methods as needed.

Food and Drug Administration (FDA) is the responsible bureau of food, medicine, biological medical products, blood products, medical instruments, radiation emitting instruments, veterinary instruments and cosmetics in United States Ministry of Health.

Medical devices and medical products are classified according to their potential risk situation due to their design and manufacture, and the degree of danger to human health. FDA categorizes medical devices and medical products to three categories, like class III: high risk products, class II: medium risk products and class I: low risk products.

The European Union has directives for medical devices. These directives including the technical requirements and responsibilities that must be provided during the design and production phase of medical devices before they released:

- •Directive 90/385/EEC regarding active implantable medical devices;
- •Directive 93/42/EEC regarding medical devices;
- •Directive 98/79/EEC regarding in-vitro diagnostic medical devices.

Institute of Physics and Engineering in Medicine (IPEM), aims its members to apply the progress in the field of physics and engineering to the field of medicine and biology, and to increase the knowledge of people in medicine and biology with their education and present these to the use of interested persons.

The European Metrology Research Program (EMRP) and the European Metrology Programme for Innovation and Research (EMPIR) which is being executed under Article 169 of the European Commission include Common Research Projects which have been launched in various "metrology in health" fields between 2007 and 2011. Projects have been conducted 4-5 NMIs/DIs and unfunded partners with a budget of approximately 3 Million Euro in three years.

In the first call of health, 7 projects were supported and completed. The projects completed with the participation of NMIs and DIs are listed below:

- •Breath analysis as a diagnostic tool for early disease detection;
- •Metrology on a cellular scale for regenerative medicine;
- •Increasing cancer treatment efficiency using 3D brachytherapy;
- •External Beam Cancer Therapy;
- •Traceable measurements for biospecies and ion activity in clinical chemistry;
- •Traceability of Complex Biomolecules and Biomarkers in Diagnostics - Effecting Measurement Comparability in Clinical Medicine.

In the second EMRP call of metrology in health, 11 projects were funded and completed:

- •Metrology for a universal ear simulator and the perception of non-audible sound;
- •Metrological characterization of micro-vesicles from body fluids as non-invasive diagnostic biomarkers;
- •Dosimetry for ultrasound therapy;
- •Metrology for the characterization of biomolecular interfaces for diagnostic devices;
- •Metrology for metalloproteins;
- •Metrology for next-generation safety standards and equipment in MRI;
- •Metrology for drug delivery;
- •Metrology for monitoring infectious diseases, antimicrobial resistance, and harmful micro-organisms;

•Metrology for radiotherapy using complex radiation fields;

•Metrology for biomolecular origin of disease;

•Metrology for molecular radiotherapy.

According to success in EMRP programme, a new programme EMPIR calls, launched between 2014 and 2020. Budgets were approximately 2 Million Euro in three years with the partnerships of external partners from universities, research institutions and private companies. 9 metrology in health projects have been initiated already (EURAMET Newsletter Issues 1-12), namely:

•Quantitative measurement and imaging of drug-uptake by bacteria with antimicrobial resistance;

•Role of metals and metal containing biomolecules in neurodegenerative diseases such as Alzheimer's disease;

•Metrology for modern hearing assessment and protecting public health from emerging noise sources;

•Innovative measurements for improved diagnosis and management of neurodegenerative diseases;

•Metrology for multi-modality imaging of impaired tissue perfusion;

•Metrology for clinical implementation of dosimetry in molecular radiotherapy;

•Novel materials and methods for the detection, traceable monitoring and evaluation of antimicrobial resistance;

•Metrology for MR guided radiotherapy;

•Metrology for additively manufactured medical implants.

In response "Metrology for Health" ranks high on the EURAMET agenda. A dedicated Task Group was established in 2013 in order to develop a strategy on how metrological R&D should evolve in the context of EMPIR and the European Union's HORIZON2020. The main objective of the Task Group is to organize the scientific and technical collaboration in health related fields among EURAMET members and associates and to form a coherent and comprehensive approach on metrology for health. In more detail the group's terms of reference ask to:

•Coordinate and to complement the work of EURAMET's Technical Committees in metrology for health;

•Liaise with the Joint Committee for Traceability in Laboratory Medicine (JCTLM) and other groups working in this field;

•Support and act for the development of standards, measurement methods and measurement structures;

•Develop the Strategic Research Agenda for EMPIR;

•Propose research topics for joint research projects and to elaborate road maps for future research and development;

•Disseminate expertise and knowledge on metrology for health through seminars, guides and conferences.

The OIML (The International Organization of Legal Metrology) main goal is harmonization of regulations and metrological controls which are applied by national metrology institutions of member states or related organizations (Ferreira, 2011). There are two basic categories of OIML publications.

OIML Draft Recommendations and Documents are prepared by technical committees or subcommittees which are created by Member States. One of these technical committees is the technical committee of TC 18 Medical Measurement Devices. In this technical committee, subcommittees were formed in different subjects:

•TC 18 – Medical measuring instruments;

•TC 18/SC 1: Blood pressure instruments;

•TC 18/SC 2: Medical thermometers;

•TC 18/SC 4: Bio-electrical instruments;

•TC 18/SC 5: Measuring instruments for medical laboratories.

3. Measurement Parameters and Traceability in Medical Device

A traceability scheme is given in Figure 4 for medical measurements. For example, blood pressure measuring devices (that are individual or engaged to a patient monitor) are calibrated by using patient simulator device that are calibrated routinely in NMSs or DIs. Establishment of similar links to calibrators and so SI/derived SI units maintain the accuracy in the medical measurements.

Patient simulator, defibrillator/pacer, infusion pump simulator, electrosurgery analyzer, electrical safety analyzer, pulse oximeter analyzer and gas flow analyzer are mostly used medical device calibrators. Parameters, measurement ranges and traceabilities were given in Table 1-Table 7 (1).

For example, calibration of infusion pump analyzer has realized according to manufacturer specifications. Before the calibration process, the device flow channels are cleaned by flowing distilled and degassed water with soft detergent by connecting the Syringe3 directly into channels in order to remove any remaining dirt and chemicals residues. This process prevents the formation of unwanted bubbles during calibration. All those connections are completed and filled with water in order to avoid bubbles in the water during calibration as seen in Figure 3. The cleaning process repeated few times until no bubbles left.

Than calibration process is realized by the following:

•First, some amount of distilled and degassed water is injected from reservoir to fill in the Syringe3 as in the ways of red arrows while valves V1, V2 and V3 closed and valve V4 is open;

•Then V4 is closed and V1 opened. The water is sent to calibrated Syringes 2 (and Syringes3 for calibration of 2 channels at the same time) as in the ways of blue arrows until they are full;

•Infusion device has been set to desired flow rates and volumes between 75 ml/hr and 830 ml/hr and at 25 ml and 45 ml volumes according to manufacturer specifications;

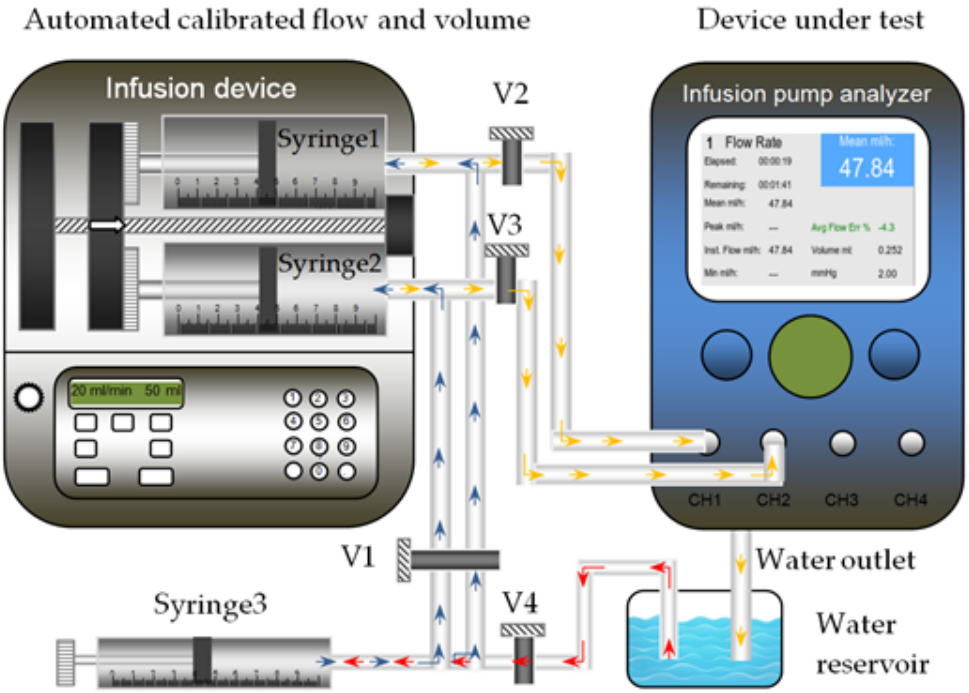
•The process started automatically for each set of measurements. The water is flowed through the channels of infusion pump analyzer as in the ways of yellow arrows while V1, V4 are closed and V2, V3 are opened;

•Pressure tests are realized between 5 PSI and 40 PSI;

•Solenoid valve release test are conducted and value is measured;

•All values are measured and recorded for each measurement at each channel and compared to tolerances and manufacturer specifications.

Figure 3 - Infusion pump analyzer calibration system.



4. Conclusions

Medical metrology concept has been investigated in detail in this paper. Accuracy in medical measurements satisfies the more accurate diagnosis and treatment and less expenditure in health expenses. Even though metrology is an internationally linked activity, current situation shows that each NMI has some kind of studies in medical metrology field that are not harmonized between each other. International organizations (i.e. JCI, ECRI) have individual guidelines for quality control of medical devices but don't directly aims to metrology traceability.

NMIs or DIs have to have a collaborative and multidisciplinary study in order to establish the traceability and equivalence of medical measurements in all over the world. Another important issue becomes to educate and train medical people who are engaged in use of medical devices and measurements. Scheduled works in the field of medical metrology by NMI or DI can be listed as follows:

•Establishment of traceability in diagnostic and therapeutic devices in the field of health;

•Investigation of required laboratory infrastructure for validation of measurement method used to verifying measurements, analysis and tests in clinical measurements;

•Production of reference (or quality control) materials used in clinical measurements;

•Creation of metrology awareness for universities, calibration laboratories and societies which are interested with this subject;

•Intercomparisons and competence tests for the purpose of ensuring measurement unity between laboratories (i.e. NMI, DI or secondary laboratories) in the country and in the world.

References

Ainsley, C.G., Yeager, C.M.(2014), Practical considerations in the calibration of CT scanners for proton therapy, *J Appl Clin Med Phys.* May 8;15(3):4721. doi: 10.1120/jacmp.v15i3.4721, PMID: 24892347.

Baura, G. D. (2012), *Medical Device Technologies: A Systems Based Overview Using Engineering,* Academic Press, 2012, 512 pages.

Bošnjaković, A., Džemić, Z.(2017), Legal Metrology: Medical Devices, *CMBEBIH 2017* pp 583-588, © Springer Nature Singapore Pte Ltd. 2017 583, A. Badnjevic (ed.), CMBEBIH 2017, *IFMBE Proceedings* 62, DOI: 10.1007/978-981-10-4166-2_88.

Chiao, J. C., Goldman, J.M., Heck, D.A., Kazanzides, P., Peine, W.J., Stiehl, J.B., Yen, D., Dagalakis, N.G.(2008), Metrology and Standards Needs for Some Categories of Medical Devices, *J Res Natl Inst Stand Technol.* Mar-Apr; 113(2): 121–129, PMID: 27096115.

Feldman, M.D., Petersen, A.J., Karliner, L.S., Tice, J.A., (2007). Who is responsible for evaluating the safety and effectiveness of medical devices? The role of independent technology assessment. *J Gen Intern Med.* 23(Suppl 1):57–63.

Ferreira, M. (2011), The role of metrology in medical devices, *OIML Bulletin Volume* L II, Number 4.

Ferreira, M. (2011), The role of metrology in the field of medical devices. *International Journal of Metrology and Quality Engineering* **2(02):**135 – 140, DOI10.1051/ijmqe/2011101.

Food and Drug Administration - FDA, http://www.fda.gov/MedicalDevices/default.htm.

Hermanek, P., Carmignato, S.(2017), Establishment of metrological traceability in porosity measurements by x-ray computed tomography, *Conference: Developments in X-Ray Tomography* XI-DOI10.1117/12.2276942

Ittermann, B., Bottauscio, O., Hand, J. Pooter, J., Prez, L., Rabus, H., Seifert, F., Szymanowski, H., Weidemann, G., Zilberti, L.(2015), Metrology for MRI safety, *17th International Congress of Metrology* doi. org/10.1051/metrology/20150009001.

Karaböce, B., Gülmez, Y., Akgöz, M., Kaykısızlı, H., Yalçınkaya, B., Dorosinskiy, L.(2015). Medical metrology studies at TÜBİTAK UME, *17th International Congress of Metrology,* DOI: 10.1051/20150 011.

Karaböce, B. (2016), Metrological approach in medical measurements. *2nd International Conference on Medical Physics and Biophysics.* Barcelona, Spain.

Monteiro, E.C, Leon, L.F.(2015), Metrological Reliability of Medical Devices, *J. Phys.: Conf. Ser.* 2015, 588 012032.

Schreyögg, J., Bäumler, M., Busse, R. (2000), Balancing adoption and affordability of medical devices in Europe. *Health Policy* 92(2-3):218-24, DOI10.1016/j.healthpol.2009.03.016.

Semerjian, H.G., Watters, R.L.(2000), Impact of Measurement and Standards Infrastructure on the National Economy and International Trade. *Measurement* 27, 179–196.

Shirmohammadi, S., Barbé, K., Grimaldi, D., Rapuano, S. and Grassini, S. (2016). Instrumentation and Measurement in Medical, Biomedical, and Healthcare Systems. *IEEE Instrumentation & Measurement Magazine,* pp, 6-12.

Squara, P., Imhoff, M., Cecconi, M. (2015), Metrology in Medicine: From Measurements to Decision, with Specific Reference to Anesthesia and Intensive Care, *Anesth Analg.* **120(1):** 66–75. PMID: 25625255.

Tooker, J. (2005),The Importance of Measuring Quality and Performance in Healthcare, *MedGenMed*. **7(2):** 49, PMID: 16369427.

Zaitseva, E.(2017), *Importance Measures in Reliability Analysis of Healthcare System, Human – Computer Systems Interaction: Backgrounds and Applications 2, Part of the Advances in Intelligent and Soft Computing book series (AINSC, volume 98),* 2017 Springer International Publishing AG. Part of Springer Nature.

Zimmerman, M. (2014),The Importance of Measuring Outcomes in Clinical Practice. *US Psychiatric & Mental Health Congress.* Career, Major Depressive Disorder, Volume: 31Issue: 10.

Authors Profiles:

Baki Karaböce has been working as a lead researcher in the field of acoustics, vibration, ultrasonic, electrical, mechanical and medical measurements at TÜBİTAK UME (National Metrology Institute) since 1993. He is the founder of the Medical Metrology Research Laboratory and responsible for the metrology in health at UME. He has participated and successfully completed several projects in various periods in 7th Framework and Horizon 2020 program. He has prepared number of book sections, refereeing tasks in congresses and journals, and numerous scientific publications at national and international level. He has served as referee and adviser in TUBITAK TARAL (national funding) projects. He has participated and organized national and international congresses as invited speakers, organizers of sessions and presidencies in congresses. He has chaired the Turkish Acoustical Society. As a proposer and member of the Health Working Group set up at EURAMET (European Metrological Centers Association), he is actively involved in the organization, programming and strategy preparation of all projects. Baki Karaböce is Turkish representative of the EURAMET TC-AUV (Acoustic, Ultrasonic and Vibration) Ultrasonic section. He is an active member of IEC (International Electrotechnical Commission)’s TC87 Ultrasonics group.

Statistical methods for a comparative study on Health Metrology

Maria do Céu Ferreira
mcferreira@ipq.pt
IPQ-Instituto Português da Qualidade

Abstract:

To evaluate the role of metrology in health care facilities, a study was performed among Portuguese hospitals to identify and analyze issues concerning the concept of metrological traceability. The discussion is structured around a dataset obtained from a survey questionnaire covering 44 public and private Portuguese hospitals. The principal results of our analysis include identifying some key indicators that apply to certified/accredited hospitals. Test reports and calibration certificates were the main metrological traceability documents required by the hospitals. The acquisition of medical measuring instruments with conformity reports depended on the type of hospital. Maintenance was considered the most relevant issue for private hospitals, whereas for the public hospitals this issue is less important than trust in the supplier. For public hospitals, purchase price was the most important requirement. For hospitals that acquire calibrated measuring instruments, only 50 % perform in-house calibration. We conclude that trust in supplier is associated with a lower overall budget for maintenance operations. The protocol established by hospitals to acquire measurement instruments does not reflect metrological concerns and the relationship between maintenance and metrological operation is not well established. Thus, we conclude that metrological traceability is largely absent in the daily considerations of health care professionals and facilities.

Keywords: Metrology; Metrological Traceability; Measuring instruments, Health, Quality assurance

1. Introduction

Measurement, as a field of scientific and technical activity, encompasses a set of activities aiming at determining values for quantities of interest. It has shown a remarkable evolution through the centuries (Kind & Lubbig, 2003). The scientific area that is concerned with these activities is called Metrology (JCGM 200, 2012), and it covers a wide range of key activities in all sectors of society, with a large impact on the economy (Hratch et al., 2000). Metrological traceability, measurement uncertainty, and method verification/validation are the main elements that support the capability of comparing measurements. Measurement uncertainty plays in turn an important role in assessing conformity with limits as well as in the comparison of measurement data (Golze, 2003).

Measurements and measuring instruments are present in day-to-day life and are fundamental in the prevention, diagnosis and treatment of diseases. Healthcare professionals assume that these measurement results are accurate, reliable and comparable over time. This assumption implies that they believe measurement results are metrologically traceable and that the methods and procedures with which they are obtained have known performance characteristics, often expressed in terms of repeatability and reproducibility (Ashford, 2010). Nevertheless, there should be awareness that errors and uncertainties in the use of measuring instruments, methods and procedures could lead to unsuitable clinical procedures, wrong diagnoses, or inadequate treatments.

Developments in clinical measurements constitute a good example for the improvements achieved by paying better attention to metrological traceability in laboratory medicine. The Joint Committee for Traceability in Laboratory Medicine (JCTLM), hosted by the International Bureau of Weights and Measures (BIPM), is the first international committee focusing on metrology in clinical chemistry and is a platform for the harmonization and dissemination of good practices in this field (Armbruster & Richard, 2007). Under the scope of the European Medical Devices Directives, new responsibilities and technical requirements for the measuring instruments were defined. Although this directive covers a wide range of instruments, the legal framework still allows each member state to consider complementary measures to protect public health and its citizens. At the same time, no further regulated control (according to EU policy) exists for these medical devices and instruments after they are placed on the market and put into service (Ferreira, 2011).

As the awareness of the benefits for having traceable measurement results increases, some researchers and technical experts (in metrology and medicine), have increased their efforts and interest for developing good and best measurement practice, based on metrological principles. Several papers were recently published focusing on specific types of measuring instruments used by health professionals. Some authors highlighted the key role of metrological traceability of sphygmomanometers as a tool for reducing health care costs and improving patient's care (A'Court et al.,2011). Similarly, other researchers pointed out the need for traceability in clinical temperature results (Chung & Chen,2010). With respect to ophthalmic diseases, the evaluation of metrological concerns related to tonometers is under investigation (Choudhari et al., 2009). Protocols and good practices for using measuring instruments for pulmonary and respiratory diseases have been developed (Jensen et al., 2007).

Notwithstanding the progress made in the aforementioned work, much remains to be done to improve measurement practices in health care. In some fields of activity, the jurisdiction action players in health care have not done enough to assure the metrological traceability, and by implication the reliability, of health-related measurements. Furthermore, the growing concerns about quality assurance in health care facilities bring new areas of metrological opportunity. The lack of awareness of health professionals regarding metrology is a topic requiring further evaluation. Moreover, it is also considered that the interoperability between metrology, quality assurance, and health care is not contextualized within the current legal European framework.

2. Method

A survey was developed that took into account the literature on this subject and the current framework of national health care facilities. The survey design and the data analysis were based on a multi method design as a qualitative (content analysis) and quantitative research methodology (Brannem, 2005). The analysis methods included inferential statistical analysis, nonparametric tests, Multiple Correspondence Analysis (MCA) and Exploratory Factor Analysis (EFA) (Maroco, 2007). In order to compare the relative likelihood between cross-variable outcomes, it was also used an Odds Ratio (OR, θ) evaluation.

Data was collected using a survey structured as a questionnaire.

Survey characterization

The Portuguese hospitals were the target of this study and included private and public hospitals. Under the scope of this study, public hospital is an official hospital under the supervision of the ministry of health. The survey was focused on the hospitals included in the database from Portuguese Directorate-General of Health (DGS) available at that time (DGS,2008), without comparable information about private hospitals. Therefore, the population of private hospitals was identified by in situ regional search.The survey was designed and developed based on the analysis of the National and European metrological requirements for health care facilities. The research hypotheses and the development of the measure survey were performed using an exploratory qualitative approach without a previous similar validated survey.

From May 2011 to September 2011, 122 questionnaires were sent by e-mail to Portuguese hospitals. From those, 99 were addressed to public hospitals, including those located on Portuguese islands (overseas territory); 35 responses (35 %) were received. The remaining 23 were sent to private hospitals; 9 responses (39 %) were received. The overall response rate was 36 %, which was considered sufficient for continuing with the study.

Table 1- Questionnaire outline.

Item	Variable	Question type/Measurement scale
1-Hospital characterization	1.1 Classification 1.2 Geographical localization 1.3 Habitants 1.4 Number of clients	Open/Nominal[a]
2-Type of Hospital (quality assurance)	2.1 Certified hospital 2.2 Accredited hospital 2.3 Some units qualified	Closed/Nominal
3-Quality Management Systems	3.1 Identify the standard references 3.2 Entity responsible for the qualification 3.3 Qualification date 3.4 Renewal Qualification assessment	Open/Nominal
4-The measuring instruments acquisition	4.1 Price 4.2 Trust in the supplier 4.3 Requirements of ISO 13485 [c] 4.4 Maintenance conditions 4.5 History of the equipment [d] 4.6 Requirements of the Notified Body	Closed/*Likert* [b]
5-Metrological terms of reference applied to measuring instruments documentation	5.1 Calibration certificate 5.2 Verification certificate 5.3 Test report 5.4 Conformity report (product) 5.5 Without certificate	Closed/Nominal
6-Interface health professional *vs* measuring instrument	6.1 Metrological knowledge for measuring instruments 6.2 Interpretation of certificates vs test reports 6.3 Relation adverse event vs instrument error 6.4 Relation instrument error vs indication results	Closed/*Likert*
7-Budget for medical equipment with measuring functions	7.1 Preventive maintenance 7.2 Corrective maintenance 7.3 Calibration 7.4 Verification	Closed/*Likert*
8-Metrological factors as a transversal tool in health care facilities	8.1 Relevance of the traceability of medical instruments 8.2 Metrological Legislation 8.3 External calibration 8.4 *In house* calibrations	Closed/Nominal
9-Training in metrology	9.1 Identify health professionals with metrology competences.	Closed/Nominal
10-The role of measures in the national economy	10. Identify the level of importance concerns the metrological traceability in the health economy.	Closed/Likert

[a] Responses to open questions were transformed into a nominal measurement scale.

[b] A *Likert* scale is an ordinal scale that uses a set of responses ordered so that one response is greater than another. This scale can be used for any question that has 3 or more possible options.

[c] Requirements concerning the fulfillment of medical devices manufactures under the scope of Portuguese policy.

[d] The "history of the equipment" means the evaluation of the test records across time.

Pretesting was conducted to assure the survey's validity. Apart from slight changes, the questions were maintaine da fter the pretesting. Data analysis was performed using Statistical Package for Social Sciences (IBM SPSS), version 19. The questionnaire achieved a Cronbach alpha of 0.74, confirming its acceptable internal consistency. In a first stage, univariate analysis was used to characterize the trend of the responses in each item measure score. Descriptive statistics were applied by computing the average, standard deviation, minimum and maximum values and frequency response. The second stage concerned the correlation between variables, including a bivariate statistical analysis from the dataset. The third and final stage included statistical inference, nonparametric tests, MCA and EFA evaluation.

3. Results

Results were obtained from analysis of the following main topics as listed in Table 1:

- •Quality systems of survey respondents (items 2 and 3);
- •Procedure for acquiring equipment vs metrology (items 2, 4 and 5);
- •Metrology vs risk management vs patient safety (item 6);
- •Budget for metrological operations and maintenance procedures (item 7);
- •Metrological factors as a transversal tool in health care facilities (item 2 and 8);
- •Quality systems of survey respondents.

The Quality Management System (QMS) overview results indicated that 53 % of the respondents had some units certified/accredited as a partial qualification. As a whole recognition of their QMS, accredited hospitals represent 25 % of the respondents whereas 23 % were certified.

More than 50 % of the respondents were qualified according to international standards, namely ISO 15189, ISO 22000 and ISO 9001. National accreditation and certification bodies were preferred as the qualification agency by both private and public hospitals.

Procedure for acquiring equipment versus metrology

Considering the policy requirements to purchase medical equipment, some relevant factors that could influence equipment acquisition was assessed. For this purpose, EFA analysis was performed with orthogonal Varimax Rotation (Maroco, 2007). Sampling adequacy was confirmed by the KMO (Kaiser-Meyer-Olkin) and the Bartlett tests. The computed outcomes are listed in Table 2. Three Varimax factors have eigenvalues greater than 1.0 and are appropriate for describing latent structure. All three of these factors contain scores with absolute values greater than 0.707 (the square-root of 0.5) which indicates that the latent structure can be identified.

The variables are thus grouped by the following latent factors:

- •technical conditions factor (1st EFA factor), consisting of maintenance conditions and the history of the equipment;
- •legal requirements (2nd factor), including the compliance of requirement procedures as well as the hospitals' requirements for the notified bodies;
- •acquisition factor (3rd factor), including the price and the trust in supplier.

Table 2- Exploratory Factorial Analyses (EFA) of the variables listed for item 4 of the questionnaire.

	Factor			
Variable	**1**	**2**	**3**	**Communality**[a]
Price	-0.21	0.00	**0.83**	0.73
Trust in supplier	0.29	0.05	**0.75**	0.65
Requirements of ISO 13485	-0.07	**0.89**	0.14	0.82
Maintenance conditions	**0.80**	0.27	0.13	0.73
History of the equipment	**0.77**	-0.05	-0.05	0.59
Requirements of the Notified Body	0.44	**0.71**	-0.13	0.71
Eigenvalue[b]	1.93	1.30	1.01	

a Communality measures the variance explained by all factors and must be considered when it is greater than 0.5.

b The eigenvalue measures the relative variance explained by each factor and must be considered when it is greater than 1.0.

By cross analyses, these variables were determined to be related the type of hospital. Therefore, the items scores were transformed onto a 5 point Likert scale by an interval range from unimportant to extremely important. These results are presented in Figures 1 and 2 as percentages. For example, private hospitals (Figure 1) considered maintenance to be the most relevant issue (89 % considered this requirement extremely important) whereas public hospitals (Figure 2) considered it to be less important (just 37 % of respondents saw it extremely important). For public hospitals, price was the most important factor (51 %) which represents a strong restriction for the enforcement of metrological traceability.

Figure 1- Private hospitals: percentage of Likert levels attributed to the given indicators.

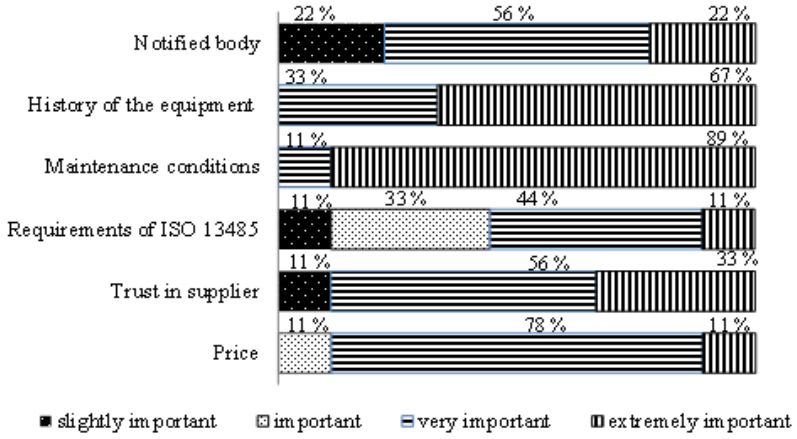
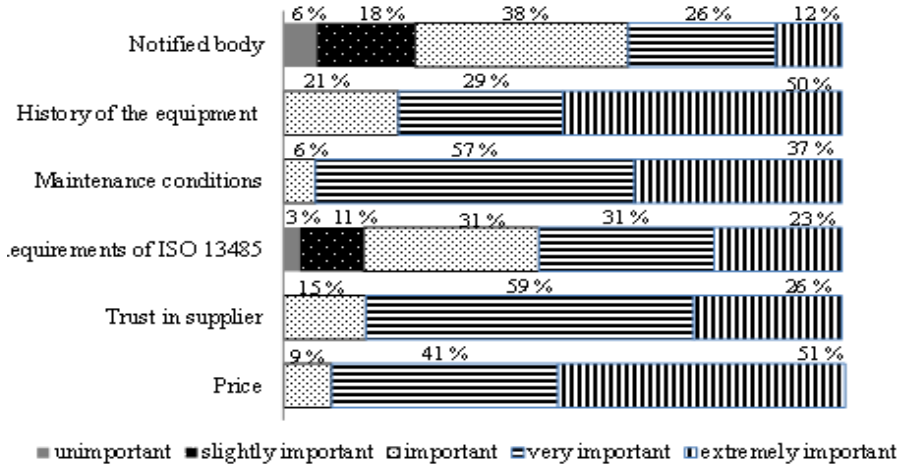


Figure 2- Public hospitals: percentage of Likert levels attributed to indicators given.



On the context of the purchase policy (in terms of equipment acquisition), metrological documentation that is included on reference terms was another aspect that was analyzed. In order to analyze the relations between the nominal variables, MCA was applied. MCA is a multivariate method that allows exploring the principal components of a dataset by transforming the variables into dimensions through orthogonal projection. Similarly to EFA, the retention criteria were eigenvalues greater than 1.0 and principal component values greater than 0.5. The MCA results are shown in Table 3.

The three retained principal dimensions explain almost 75 % of the data variability. The first dimension represents the verification certificate and test report whereas the second and third represent the calibration certificate and conformity report. Respondents who have no preference for calibration certificates are present in the 1st dimension, which are replaced by verification certificates and test reports. The opposite is true for the 2nd dimension. The 3rd dimension represents hospitals that receive conformity (evaluation) reports.

The chi square value c2 ($\alpha = 0.05$) was computed to analyze the relationship between the variables described in Table 3 (item 5 of the questionnaire) and the types of hospitals. Since the null hypothesis H0 is rejected ($c2 = 6.92$; $p = 0.008$), we conclude that the acquisition of medical measuring instruments with conformity report depends on the type of hospital. This situation was also assessed by the Odds Ratio (OR, θ) measure, a comparison of the odds of observing an outcome given exposure to a presumptive cause relative to the odds of the same outcome in the absence of exposure to that cause (Morris & Gardner, 1988). The OR for private hospitals choosing to use conformity reports is only 0.65, the OR for public hospitals is a quite high 8.0.

Table 3 - Survey MCA applied to measuring instruments documentation (item 5 of the questionnaire Table 1).

Variables	Principal Dimensions		
	1	2	3
Calibration certificate	0.01	**0.71**	0.098
Verification certificate	**0.52**	0.02	0.08
Test report	**0.58**	0.00	0.09
Conformity report (product)	0.29	0.03	**0.51**
Without certificate	0.00	0.46	0.36
Eigenvalue	1.40	1.22	1.11
Variance, %	28.0	24.3	22.3

The relation between hospitals that acquire equipment with calibration certificates and its relevance with respect to i) metrological knowledge and ii) technical competence for certificates' analyses was also analyzed. The results indicated that hospitals with certified equipment (calibrated measuring instruments) consider relevant their own certified interpretation (50 %). Furthermore, 46 % expressed their interest in metrological knowledge applied to clinical practice.

Metrology vs risk management vs patient safety

Another issue evaluated focused on the key role of metrology in risk management and patient safety. Two different principal dimensions were identified by MCA; they are shown in Table 4. The 1st dimension can be characterized by the association between the need to interpret metrological certificates of equipment and the reflection of this interpretation on the errors generated by the measuring instruments. Considering that these errors can be widespread in diagnostic and therapeutic procedures, extrapolation of this interpretation for clinical practice (whose influence on adverse effects can be significant) is also present in this dimension. On one hand, the metrological knowledge that health professionals must acquire for their measuring instruments is very important and can influence the evaluation of measurement error as well as its outcomes. On the other hand, the high score of i) interpretation of certificates vs tests reports and ii) adverse event vs instrument error show relevant concerns around patient safety. We conclude that the risk related to the use of medical equipment and patient safety can be considered exogenous latent variables with special relevance in health metrology.

Table 4- Survey MCA applied to the relationship health professional vs measuring instrument (item 6 of the questionnaire Table 1).

Variables	Principal Dimensions	
	1	2
Metrological knowledge for measuring instruments	0.48	**0.64**
Interpretation of certificates *vs* tests reports	**0.92**	-0.06
Relation adverse event *vs* error of instrument	**0.92**	-0.07
Relation error of instrument *vs* indication results	0.23	**-0.84**
Eigenvalue	1.19	1.13
Variance, %	49.6	28.3

Budget for metrological operations and maintenance procedures

Considering the key role of budget management, we assessed the budget outcomes from hospitals related to metrological operations and maintenance procedures, notably preventive and corrective maintenance, calibration, and verification of measuring instruments. These items were scored on a 5 point Likert scale by an interval range from less than 10 % to more than 30 % of the hospitals' overall budget allocation. Regarding maintenance operations, corrective maintenance is the operation selected by 14 % of hospitals with a budget more than 30 % whereas 75 % of hospitals allocated less than 10 % of their overall budget. That means 33 hospitals devoted less than 10 % of their budget to maintenance operations (corrective and preventive) 30 of which were public hospitals. Under this subject, the public hospitals included in this study had up to 1100 individual measuring instruments. However, only 2.3 % of the respondents reported more than forty calibrated instruments per calibration period (defined by each hospital). Therefore, one can assume the budget values applied by hospitals represent a worrying indicator and a poor predictor for the metrology in the health services.

In Table 5 results of MCA analyses for this dataset are listed. Corrective maintenance, calibration, and verification are included in the same principal dimension, being all with score close to the 1.0 saturation value. Preventive maintenance has the highest score on the other dimension. The Cronbach alpha of 0.85 for this analysis indicates an acceptable internal consistency.

Table 5 - Survey MCA applied to the budget for medical equipment with measuring functions (item 7 of the questionnaire Table 1).

Variables	Principal Dimensions	
	1	2
Preventive maintenance	0.54	**0.97**
Corrective maintenance	**0.85**	0.02
Calibrations	**0.93**	-0.31
Verifications	**0.94**	-0.03
Eigenvalue	2.75	1.12
Variance, %	75.3	23.4

Metrological factors in health care facilities

Metrological considerations are distributed throughout health care decision processes, which led to its inclusion in several questions. The relevant variables are those of item 8 of the questionnaire (Table 1). Since the responses to these questions were not in harmony (Cronbach alpha of 0.25), a binomial test was applied by which the null hypothesis was rejected for metrological legislation. That is, the ratio of hospitals that consider legislation in the field of traceability of medical devices as important is statistically similar to the proportion of those who do not consider it to

be relevant. In order to validate this inference, a MCA was applied and two principal components (dimensions) were retained. In Table 6 the results of this analysis are presented. One dimension is dominated by metrological legislation. The other dimension is dominated by in house calibrations and can be interpreted as concern measurement traceability.

Table 6 - Survey MCA applied to the metrological factors in health care facilities (item 8 of the questionnaire Table 1).

Variables [a]	Principal Dimensions	
	1	2
Relevance of the traceability of medical instruments	0.32	0.34
Metrological Legislation	**0.64**	0.01
External calibrations	0.24	0.31
In house calibrations	0.03	**0.59**
Eigenvalue	1.23	1.15
Variance, %	30.7	28.7

Among the respondents that considered traceability to be relevant, 93 % carried out external calibration and 64 % in house calibrations, respectively. These procedures (external and in house calibrations) are practiced by only 31 % of the respondents.

The OR applied to in house calibrations by type of hospital was also evaluated. The odds ratio of private hospitals performing in house calibrations is 1.5 higher than for public hospitals.

4. Discussion

A large majority of the organizations that responded to our survey had implemented a recognized quality management system where more than 50 % of the hospitals were qualified under the scope of international standards.

The purchase price of measuring instruments is the most important requirement for public hospitals. For these hospitals, there is also significant correlation between the trust in the supplier and the price of the product. This could become an even greater issue if the economic downturn continues. Our results also stressed that the high degree of confidence in the supplier is proportional to a lower overall budget for maintenance operations. That seems an unclear frontier between maintenance and metrology operations.

There is no evidence of correlation between metrological documentation (test records and metrological certificates) and the nature of the hospitals' quality management systems. Hence, despite the statistically significant results obtained for maintenance operations, we conclude that the protocol established by hospitals to acquire measuring instruments does not reflect metrological concerns.

Discovering the relationship between adverse events and the measuring instrument performance was fairly valued by the hospitals surveyed. This can be considered an indicator of metrological awareness on the part of health-care professionals. Organizations that acquire equipment with calibration certificates are those who consider an improving knowledge of metrology as important. Hence our results suggest that some health care professionals are aware of the role metrological performance of their equipment has in their clinical practice. However, in most institutions, metrological traceability is secondary to maintenance operations.

The majority of public hospitals surveyed used external calibrations without in house calibrations. Private hospitals carried out instrument calibration by themselves with technicians having had specific training in metrology.

The respondents to our survey had little knowledge of our opinions about metrological legislation. A trend for the application of metrological practices prevailed in those entities with more knowledge and awareness of metrology issues, particularly the implementation of procedures that promote metrological traceability of measurement results.

5. Conclusions

We conclude that metrological traceability is a subject that plays little role in the day-to-day life of health care professionals, although the concept is known. Our study has identified some factors the perception of which might improve the knowledge in the field of health and metrology.

Considering the key role of quality assurance and quality control in health care facilities, this research work also contributes to the introduction of metrological concepts, principles and practices in hospitals and clinical laboratories.

Within the framework of quality improvement in health care at the European level, a continuous international collaboration of experts is needed for promoting a comprehensive approach to promoting good measurement practices. This is the major challenge that should receive the attention of the metrological and the health care communities.

References

Kind, D., Lubbig, H. (2003), Metrology – the present meaning of a history term. *Metrology*. **40:** 255-257.

JCGM 200. (2012), *International Vocabulary of Metrology. Basic and General Concepts and Associated Terms.* BIPM, Sevres. URL: http://www.bipm.org/.

Hratch, G.; Semerjian, R., Watters, J. (2000), Impact of measurement and standards infrastructure on the national economy and international trade. *Measurement.* **27:**179-196.

Golze, M. (2003), Why do we need traceability and uncertainty evaluation of measurement and test results? *Accreditation Quality Assurance.***8:**539-540.

Ashford, P. (2010), *Traceability.* Cell Tissue Bank. 11:329–333. DOI 10.1007/s10561-010-9195-3.

Armbruster, D.; Richard, R. (2007),The Joint Committee for Traceability in Laboratory Medicine (JCTLM): A Global Approach to Promote the Standardization of Clinical Laboratory Test Results. *Clinical Biochemestry Review.* **28(3):** 105–114.

Ferreira, M. C. (2011), The role of metrology in the field of medical devices. *International Journal of Metrolology and Quality Engineering.* **2:**135–140. DOI: 10.1051/ijmqe/2011101.

A'Court, C.; Stevens, R.; Sanders, S. et al. (2011), Type and accuracy of sphygmomanometers in primary care: a cross-sectional observational study. *British Journal of General Practice.* 61(590):598-603. DOI: 10.3399/bjgp11X593884.

Chung, W., Chen, C. (2010), Evaluation of performance and uncertainty of infrared tympanic thermometers. *Sensors (Basel).* **10(4):**3073-89. DOI: 10.3390/s100403073.

Choudhari, S., George, R., Baskaran, M. et al. (2009), Measurement of Goldmann applanation tonometer calibration error. *Ophthalmology.* **116(1):**3-8.

Jensen, R., Teeter, J., England, R., White, H., Pickering, E.(2007), Instrument Accuracy and Reprodutibility in Measurements of Pulmonary Function. *CHEST.* 132:388-395. doi:10.1378/chest.06-1998.

Brannen, J. (2005), Mixing methods: the entry of qualitative and quantitative approaches into the research process. *International Journal of Social Research Methodology.* **8(3):**173-184.DOI: 10.1080/13645570500154642.

Maroco, j. (2007), *Statistical analysis using SPSS* (3rd ed.). Walker Editions. ISBN: 978-972-618-452-2.

Morris, J., Gardner M. (1988), Calculating confidence intervals for relative risks (odds ratios) and standardised ratios and rates. *British Medical Journal.* **296:**1313-6.

DGS (2008). *Elementos Estatísticos: Informação Geral. Direcção de Serviços de Epidemiologia e Estatísticas de Saúde.* Divisão de Estatísticas de Saúde. Lisboa. ISSN 0872-1114.

ISO 13485: 2016. *Medical devices-quality management systems – requirements for regulatory purposes.* International Organization for Standardization, Geneva, Switzerland.

ISO 15189:2012 *Medical laboratories - Requirements for quality and competence.* International Organization for Standardization, Geneva, Switzerland.

ISO 22000: 2018 *Food safety management.* International Organization for Standardization, Geneva, Switzerland.

ISO 9001:2015 *Quality management systems-requirements.* International Organization for Standardization, Geneva, Switzerland.

Authors Profiles:

Maria do Céu Ferreira holds a Ph.D. in industrial engineering from the Faculty of Science and Technology; FCT, NOVA University of Lisbon. Since 1994, she is Metrologist at the Metrology Department of the Portuguese Institute for Quality (IPQ), having developed its activities in the areas of scientific and applied metrology, managing multi-disciplinary teams related with standards and measurements. Since 2008, Maria is in charge at Legal Metrology as Quality Manager and Assessment of the Metrological operations and verification bodies. Meanwhile, she founder and is responsible for the Portuguese Health Metrology committee, hosted by IPQ. Maria is a designated member of the Metrology Board of the Portuguese Engineers order and is Professor in the Faculty of Engineering at the Lusófona University of Lisbon. Maria is the Portuguese member of the International Board of Experts of the International Organization of Legal Metrology and is an active member of OIML, WELMEC and ISO. Her research interests include all health metrological applications for ensuring traceability and new quality assured measurement methods.