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

Foundational Imperatives for Measurement with Mathematical Models

Nikolaus Bezruczko[1]

[1]Measurement and Evaluation Consulting, Chicago, USA

Abstract – A contemporary measurement problem is proliferation of mathematical models that purport to have objective measurement properties but without demonstration of meaningfulness or efficiency. This presentation emphasizes broad prevalence of this problem across nonphysical observations. A rigorous standard of practice was proposed that imposes explicit, well-established measurement principles on interpretations of mathematical measures. Principles proposed are objectivity, invariance, linearity, precision, and simplicity. Function of these principles during measurement was demonstrated during construct consolidation of scales calibrated separately for measuring health-related quality of life (HRQOL) and Spirituality, respectively, with a probabilistic measurement model. Data submitted for measurement consisted of questionnaire responses from 545 breast cancer patients to FACIT-General (28 items), which assesses general health HRQOL. Spirituality was assessed with FACIT-Sp (12 items). Separate and co-calibrated measures were compared, as well as ethnic differences (White, African-American, and Hispanic). Principal Component Analysis examined threats to dimensionality after co-calibration. In general, objectivity, invariance, and linearity demonstrated instrumental importance for establishing generality of measures, while precision and simplicity contributed to efficiency and effectiveness. Altogether, conformity to these principles established empirical foundations for asserting material grounds of a conceptual entity.

Keywords: measurement foundations, Rasch model, breast cancer HRQOL

1. INTRODUCTION

> The scientist is usually looking for invariance whether he knows it or not. Whenever he discovers a functional relationship next question follows naturally: Under what conditions does it hold? . . . The quest for invariant relations is essentially the aspiration toward generality . . . in physics, the principles that have wide applications are those we prize [1].

Stevens' insight that science advances on a range of quantitative scales (ordinal, interval, and so on) with each scale type that presents implications for meaningful measurement was a watershed in 20th century non-physical measurement. This liberation from traditional requirements of concatenation and additive scale extension encouraged a dynamic shift to mathematical representation independent of explicit units or even physical operations. Qualitative mapping and numerical representation now assert foundations for measuring previously inconceivable domains of experience. A general consequence is non-physical measurement today is associated with a range of measurement methods and variable properties of weak and strong axiomatic rigor. Less frequently recognized is Stevens' effect on *physical* measurement. Mathematical models applied in non-physical measurement are now frequently indistinguishable from physical methods, which erase deeply entrenched conceptual distinctions between physical and social phenomenology. Historical commitment to natural distinctions between physical and social secure since Aristotle in contemporary science is an anachronism. For example, complex factor analytic procedures implementing Eigen value vectors are now common not only in educational ability testing, but molecular orbital physics and quantum mechanics. Mathematical methods now commonly permeate both physical and nonphysical landscapes.

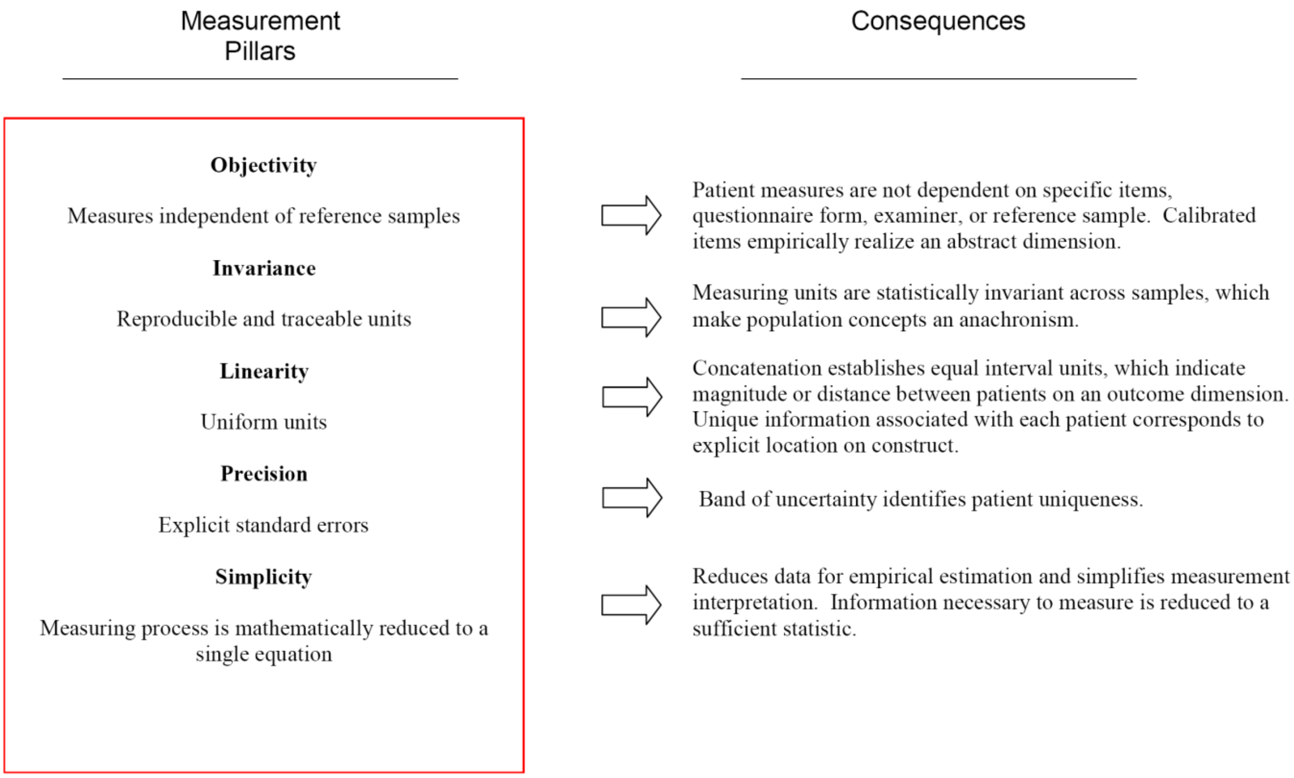
Not surprisingly, twentieth century conceptual advances have eroded traditional foundations of measurement logic and promoted practical measurement specialization. A direct consequence of this liberation from rigid, traditional measurement principles, however, is fragmentation of governing principles and enormous confusion about epistemological implications of alternative mathematical models. While measurement methods have proliferated to accommodate expanding practical needs, a general problem is coherence and validity among diverse mathematical methods and consequences of using them. A crisis of meaningfulness pervades contemporary scientific measurement [2].

Proliferation of mathematical models far exceeds capacity of contemporary conceptual frameworks to describe their function in practice or theory. Therefore, chief purpose of this presentation is to emphasize comprehensive integrative principles to re-establish universal meaning and promote cross-disciplinary, international communication. This goal is addressed in this presentation by reviewing several long standing ideas in measurement theory that could fulfil an integrative function but are currently not appreciated for coherence they offer. Consider the following:

- Invariance concepts commonly associated with factors, samples, and parameters
- Identity and uniqueness theorems that relate observations to explicit locations on finite dimensions
- Isomorphism principles that articulate between qualitative and quantitative constructs to establish congruence between sensation and number
- Abstract linear constructs that are imposed on experience by construction
- Conceptual entities that provide meaningful theoretical context for mathematical interpretation

These ideas are well-established with broad generality governing virtually all measurement practice; yet, their importance to mathematical measurement is debated, contested, and sometimes demeaned and disparaged [3]. In fact, these principles describe a fundamental process between quantitative ideas and mensuration based on a practical tradition spanning millennia [4]. Their instrumental role in constructing modern temperature and time dimensions, for example, with subsequent contributions to scientific knowledge has dramatically altered conceptions of time and space [5, 6]. A reconsideration of them now could increase coherence among mathematically diverse methods and clarify their importance for scientific theory. Figure 1 recapitulates these ideas as measurement pillars and describes their instrumental consequences for measuring patient outcomes based on self-report [7]. For example, an idealized Objectivity concept is widely accepted among measurement practitioners and has direct consequences for measuring patient attitude by implying patient responses are independent of population samples and specific questionnaire items. When Objectivity is satisfied, patient responses may be mathematically reproduced on an abstract dimension. Unfortunately, current conceptions of Objectivity are not uniformly rigorous, explicit, nor widely accepted, which weakens its influence on measurement in general. Likewise, Invariance is another widely acclaimed measurement property but without clear notions of purpose or goal. In contrast, Linearity is a technical property that is rarely asserted among mathematical methods with serious practical consequences. In contrast, Precision is a measurement property with virtually universal endorsement. Yet, Simplicity, which should also be a driver of efficiency and effectiveness, tends not to be emphasized by mathematical models.

These pillars are foundations for demonstrating continuity among mathematical measures to a larger framework defined by expanding fabric of scientific knowledge. Without this pillar foundation, linkages across scientific knowledge are incoherent and irreconcilable. A sound foundation simplifies practical measurement problems and has epistemological implications for new scientific knowledge about patients and health.



Measurement Pillars

Consequences

Objectivity

Measures independent of reference samples

Patient measures are not dependent on specific items, questionnaire form, examiner, or reference sample. Calibrated items empirically realize an abstract dimension.

Invariance

Reproducible and traceable units

Measuring units are statistically invariant across samples, which make population concepts an anachronism.

Linearity

Uniform units

Concatenation establishes equal interval units, which indicate magnitude or distance between patients on an outcome dimension. Unique information associated with each patient corresponds to explicit location on construct.

Precision

Explicit standard errors

Band of uncertainty identifies patient uniqueness.

Simplicity

Measuring process is mathematically reduced to a single equation

Reduces data for empirical estimation and simplifies measurement interpretation. Information necessary to measure is reduced to a sufficient statistic.

Figure 1. Pillars of scientific measurement have explicit consequences for rehabilitation medicine.

This presentation will demonstrate relevancy of measurement principles asserted to be pillars of scientific knowledge during empirical consolidation of two non-physical, hypothetical constructs: health-related Quality of Life (HRQOL) in oncology and Spirituality. This demonstration is conducted in context of Rasch models [8], which are currently only probabilistic models in wide use that explicitly address unidimensional measurement principles and are commonly implemented in rehabilitation medicine for quantifying patient self-reports [9].

2. METHOD

2.1. Sample

The sample is 545 patients with breast cancer diagnosis by 15 oncology specialists.

2.2. Instrumentation

The FACIT-General [10] assesses general health-related HRQOL (28 items) of cancer patients receiving treatment. The scale has demonstrated validity to discriminate patients' a) stage of disease, b) performance status rating, and c) hospitalization status, as well as sensitivity to change over time. Separate FACIT subscale measures of physical, functional, social, and emotional well-being have been validated but not integrated into an over-arching HRQOL model. In contrast, Spirituality, is typically measured separately with FACIT-Sp [11] but, in fact, would be expected to permeate the cancer HRQOL manifold.

2.3. Analysis

FACIT and Spirituality co-calibration was conducted with a Rasch model for rating scales with WINSTEPS software [12]. Conjoint probabilistic Rasch models were developed by the Danish mathematician, Georg Rasch, for transforming ordinal raw scores into objective, linear measures [4]. Rasch model transformation is mathematically based on β and δ in the following expression,

$$\Pi_{nix} = \frac{\exp \sum_{j=0}^{x} [\beta_n - (\delta_i + \tau_j)]}{\sum_{k=0}^{m} \exp \sum_{j=0}^{k} [\beta_n - (\delta_i + \tau_j)]} \quad (1)$$

where β = observations, δ = item difficulties, and τ = rating scale thresholds. Π_{nix} is the probability any item δ, will be rated X by participant β_n where X takes a value from a fixed range (j = 1, 2, 3, 4, 5, 6, . . . n), m = number of steps for an item, and k = ith step. Model prediction (P) for each item and observed ratings (O) are statistically analyzed (O-P) for significant departures from expectation with Chi-square analysis. When raw data fit a Rasch model, differences between observations have an explicit unit of measurement (logits) with axiomatic additivity independent of reference samples. Statistical estimation of model parameters implements an empirical probabilistic concatenation procedure.

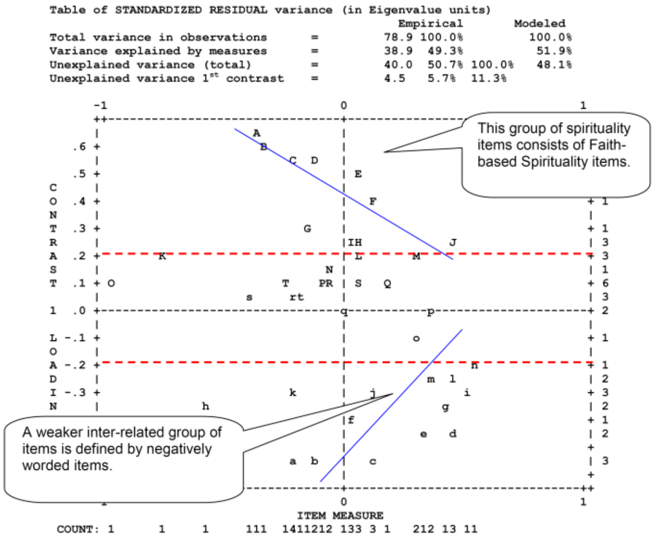
3. RESULTS

3.1 Co-calibration

HRQOL and Spiritual co-calibration was remarkably coherent. Figure 2 presents a Rasch model construct map after co-calibration, which shows items intermingle on the measurement dimension without major breaks or interruptions from low to high. Spread of items is reasonably well-centred on patients, which supports measurement of change, and item content is logically related to difficulty. Low HRQOL on this continuum is defined by items easy for cancer patients to endorse, while items higher are incrementally more difficult to endorse. Oncology patients experiencing extremely low HRQOL would be expected to endorse feelings of hopelessness and physical symptoms, while higher states of HRQOL would be consistent with endorsing Spirituality items expressing harmony and peace. In other words, results from this co-calibration are theoretically plausible, which justifies conjoint probabilistic concatenation of separate scales. Technical properties of precision and reliability also provide empirical support for measuring HRQOL and Spiritual together, which are consistent with Chi-square fit statistics. These overall results establish a first step toward consolidation.

3.2 Dimensionality threats

An issue of enormous importance when consolidating separate scales is dimensionality. Implementation of a unidimensional probabilistic model imposes strict expectations on model residuals, which were evaluated by Principle Components Analysis (PCA). Figure 3 presents PCA results, which indicates co-

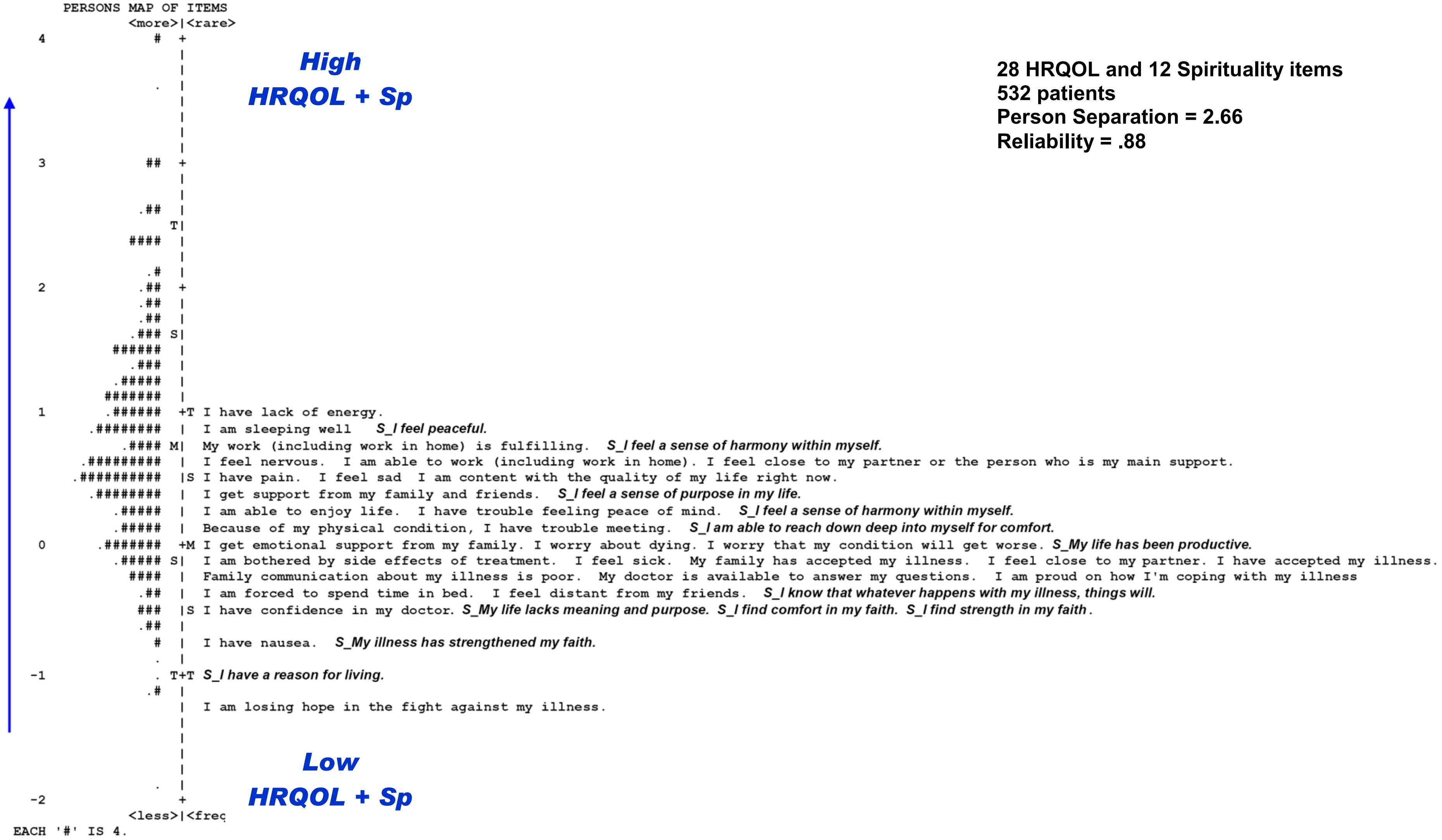


```
Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)
                                                Empirical       Modeled
Total variance in observations       =        78.9 100.0%         100.0%
Variance explained by measures       =        38.9  49.3%          51.9%
Unexplained variance (total)         =        40.0  50.7% 100.0%   48.1%
Unexplained variance 1st contrast    =        4.5   5.7%  11.3%
```

This group of spirituality items consists of Faith-based Spirituality items.

A weaker inter-related group of items is defined by negatively worded items.

Figure 3. PCA of co-calibration item residual plot. 532 patients, 40 items, 5 categories



```
      PERSONS MAP OF ITEMS
              <more>|<rare>
   4             #  +
                    |
                    |
                 .  |
                    |
                    |
                    |
                    |
   3            ##  +
                    |
                    |
               .##  |
                   T|
              ####  |
                    |
                .#  |
   2           .##  +
               .##  |
               .##  |
              .### S|
            ######  |
              .###  |
            .#####  |
           #######  |
   1       .######  +T I have lack of energy.
         .########  |  I am sleeping well    S_I feel peaceful.
             .#### M|  My work (including work in home) is fulfilling.  S_I feel a sense of harmony within myself.
        .#########  |  I feel nervous.  I am able to work (including work in home). I feel close to my partner or the person who is my main support.
       .##########  |S I have pain.  I feel sad  I am content with the quality of my life right now.
         .########  |  I get support from my family and friends.  S_I feel a sense of purpose in my life.
            .#####  |  I am able to enjoy life.  I have trouble feeling peace of mind.  S_I feel a sense of harmony within myself.
            .#####  |  Because of my physical condition, I have trouble meeting.  S_I am able to reach down deep into myself for comfort.
   0      .#######  +M I get emotional support from my family. I worry about dying. I worry that my condition will get worse. S_My life has been productive.
            .##### S|  I am bothered by side effects of treatment.  I feel sick.  My family has accepted my illness.  I feel close to my partner. I have accepted my illness.
              ####  |  Family communication about my illness is poor.  My doctor is available to answer my questions.  I am proud on how I'm coping with my illness
               .##  |  I am forced to spend time in bed.  I feel distant from my friends.  S_I know that whatever happens with my illness, things will.
               ###  |S I have confidence in my doctor. S_My life lacks meaning and purpose.  S_I find comfort in my faith.  S_I find strength in my faith .
               .##  |
                 #  |  I have nausea.   S_My illness has strengthened my faith.
                 .  |
  -1             . T+T S_I have a reason for living.
                .#  |
                    |  I am losing hope in the fight against my illness.
                    |
                    |
                    |
                    |
                 .  |
  -2                +
              <less>|<freq
EACH '#' IS 4.
```

28 HRQOL and 12 Spirituality items
532 patients
Person Separation = 2.66
Reliability = .88

Figure 2. Co-calibration variable map of FACT-G (HRHRQOL) and FACIT-Sp. FACIT-Sp (Spirituality) items are italicized.

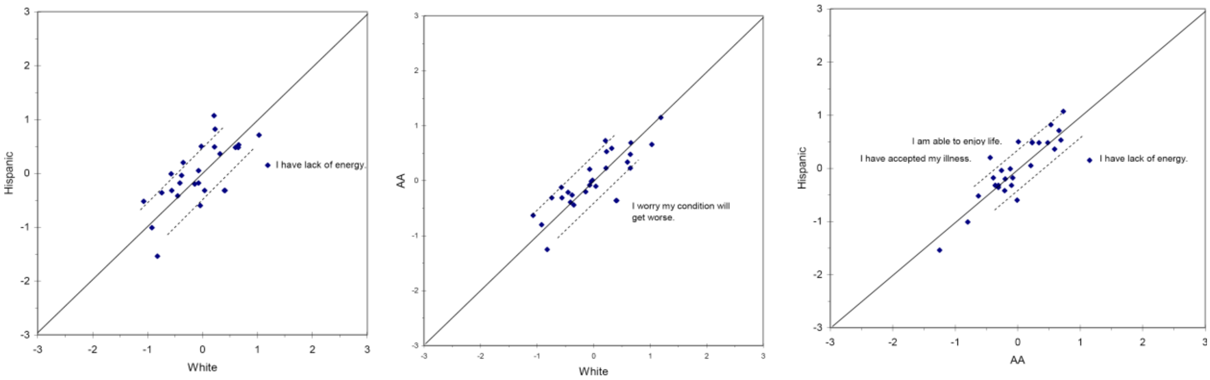


Figure 4. Bi-calibration plots of FACT-G (HRQOL, 28 items) for White, African American, and Hispanic patients.

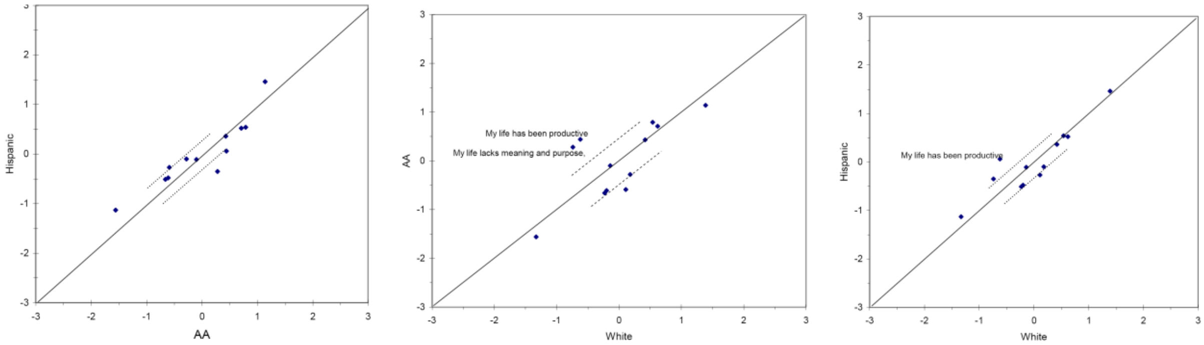


Figure 5. Bi-calibration plots of 12 FACIT-Sp (Spirituality, 12 items) for White, African American, and Hispanic patients.

calibrated dimension accounts for about 50 percent of item variance, while first PCA factor only represents five percent. Structure of dependencies in residuals, while pervasive tends to converge as item difficulties increase, which offers insight into phenomenological dynamics underlying this co-calibrated construct. Residual analysis suggests HRQOL and Spirituality influence low quality of life simultaneously but separately. Patients expressing high quality of life, however, indicate HRQOL and Spirituality are statistically indistinguishable. In other words, high quality of life for oncology patients has ordered, hierarchical relations with them.

3.3 Item stability across ethnicity

Effect of co-calibration on item stability across ethnic groups was examined first by ethnic subsets. Figures 4 and 5 show separate item calibrations were largely independent of ethnic characteristics. With only a few exceptions, items maintained an invariant structure after co-calibration.

3.4 Comparability of separate and combined scales

Finally, demonstration of item invariance leads logically to an examination of person parameters and consideration of important practical differences under conditions of (HRQOL + Spirituality) consolidation. Figure 6 presents definitive evidence that co-calibration is not only more efficient but leads statistically to identical person measures.

3.5 Consolidated scale implications

While person measures did not significantly shift after co-calibration, the band of uncertainty around measure estimates is sufficiently broad to define sig-

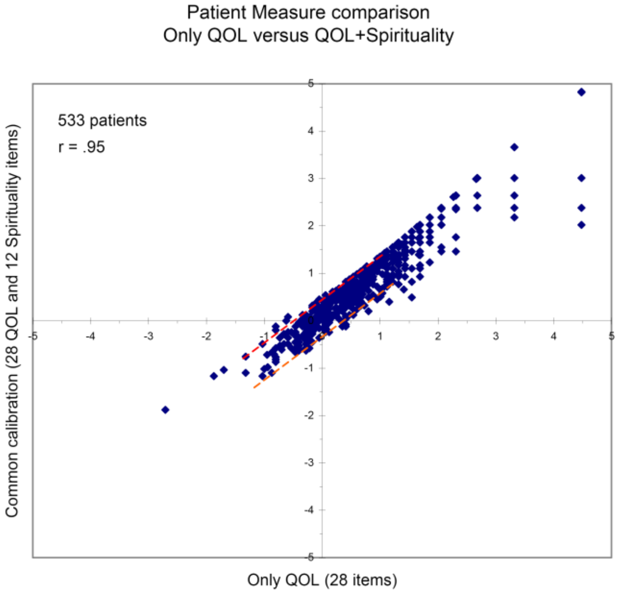
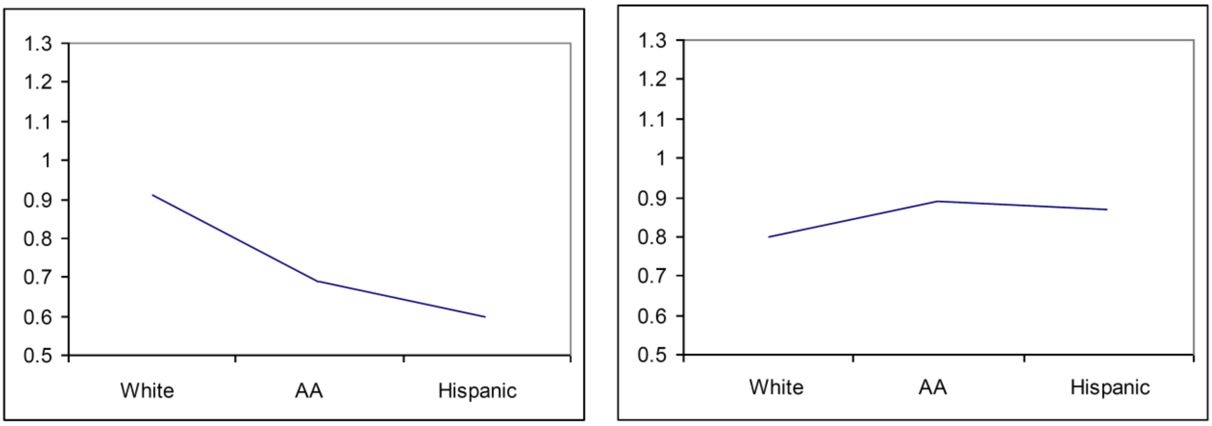


Figure 6. Patient measure comparison: Only HRQOL versus HRQOL + Spirituality items



Patient measures based only on FACT-G

Patient measures after co-calibration of FACT-G and FACIT-Sp

Figure 7. Ethnic comparison

nificantly different group means after co-calibration. Figure 7 shows Whites declined, while co-calibration measures for African Americans and Hispanics increased. These results emphasize absolute necessity for explicit, traceable units in order to conduct objective analyses of quantitative data.

4. DISCUSSION AND CONCLUSIONS

4.1 Integrative principles

The point of this presentation was to review an evaluation process that is critically dependent on mathematical models firmly grounded on measurement foundations. A purpose was to clarify consequences of specific properties for practice and interpretation. Explicit linear units, for example, ensured traceable patient measures throughout analyses above, which were necessary to verify valid links between patient self-report and eventual abstract dimensional formulation. A demonstration of probabilistic concatenation confirmed additive, linear scale units. Invariance of item parameters widely acknowledged as central to generality of measures, likewise, was dependent on uniform units estimated independently of specific empirical samples or items. Conformity to measurement logic integrated mathematical modelling with conventional steps associated with coherent assertion of material reality and conceptual entity.

Conceptual entities are a poorly understood aspect of the measuring process. They present mathematical imperatives to demonstrate parameter invariance, clarify explicit empirical boundaries, and verify isomorphism between qualitative experience and quantitative logic. When successful, this demonstration associates conceptual entities with an abstract numerical dimension, which is an important step toward linking mathematical thinking with predominant scientific conceptions. Without these imperatives, mathematical measurement remains an exercise in logic with limited value for scientific theory.

4.2. Measurement is a rationale process

Science is commitment to order and cohesion, as well as rationale understanding across vast limits of macro and micro observations. When governed by principles, objective measurement establishes observational structures that can be logically absorbed by comprehensive scientific theories. Undisciplined mathematical modelling now threatens an orderly advance of scientific knowledge by dismissing coherence dependent on measurement logic. Mathematical order pursued without sensitivity to measurement logic likely will lead to greater fragmentation and less alignment with scientific knowledge

An implied question of this research is whether measurement practice is governed by logical principles, and, if so, do they have benefits for improving meaningfulness of contemporary mathematical measurement models. Indeed, several overarching principles emerge from this research that promise to improve integration of mathematical measurement with scientific theory yet not sacrifice creativity or ingenuity. First, measurement is fundamentally an abstract linear extension, an idealization brought to empirical reality through an iterative process defined by discrete units. This articulation between abstraction and method defines a measurement procedure. When demonstration of order is not extended by concatenation, then meaningfulness symbolized by axioms has explicit implications for practice and meaning. Mathematical modelling and measurement theory face an important task of understanding and explaining their consequences for scientific theory. Long term differences may, in fact, may be irreconcilable.

4.3 Logical imperatives for mathematical measurement

If the goal of measurement is not simply quantification but meaning and understanding of numerical relations, results presented here suggest a rigorous set of measurement principles may offer reasonable guidance for any mathematical measurement model.

Prominent principles were presented above as pillars to suggest a reasonable core that probably has ultimate consequences for the façade of knowledge built on them. A goal to maintain coherent scientific knowledge in future should consider understanding this foundation.

ACKNOWLEDGEMENTS

Portions of this report were presented at 14th International Objective Measurement Workshop, New York, March 22-23, 2009 [13].

REFERENCES

[1] S. S. Stevens, "Mathematics, measurement, and psychophysics", in *Handbook of Experimental Psychology*, pp. 1-49, *Wiley,* New York, 1951.

[2] L. Narens, R.D. Luce, "Meaningfulness and invariance", In *Foundations of Measurement* by R.D. Luce, P. Suppes, D.H. Krantz, Academic Press, 1990.

[3] D. Borsboom, *Measuring the mind,* Cambridge University Press, Cambridge, UK, 2005.

[4] H. A. Klein, *Science of Measurement: A historical survey*, Simon & Schuster, NY, 1974.

[5] H. Chang, *Inventing Temperature,* Oxford University Press, NY, 2004.

[6] S. J. Gould, *Time's arrow, Time's cycle*, Harvard University Press, Cambridge, MA, 1987.

[7] N. Bezruczko, "Emergence of efficiency in health outcome measurement", *Journal of Applied Measurement*, vol. 11, pp. 197-213, 2010.

[8] G. Rasch, *Probabilistic Models for Some Intelligence and Achievement Tests*, Danish Institute for Educational Research; Copenhagen, 1960. (Reissued, University of Chicago Press, 1980.)

[9] J. C. Hobart, S. J. Cano, J. P. Zajicek, and A. J. Thompson, "Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations", *The Lancet Neurology*, vol. 6, 1094–105, 2010.

[10] D.F. Cella et al., "The Functional Assessment of Cancer Therapy", *Journal of Clinical Oncology*, vol. 11, pp. 570-579, 1993.

[11] K. Webster, D. Cella, K. Yost, "The Functional Assessment of Chronic Illness Therapy (FACIT) Measurement System: properties, applications, and interpretation", *Health and Quality of Life Outcomes*, doi:10.1186/1477-7525-1-79, 2003.

[12] J.M. Linacre, "WINSTEPS: Rasch Model Software and Manual, Chicago: Winsteps.com, 2005.

[13] N. Bezruczko and K. Perkins, "Multi-factor scale consolidation when theory is weak", Unpublished manuscript.