CRITICAL NATIONAL NEED IDEA

METROLOGICAL INFRASTRUCTURE FOR HUMAN, SOCIAL, AND NATURAL CAPITAL

Submitting Organization
www.LivingCapitalMetrics.com

Contact
William P. Fisher, Jr., Ph.D.
5252 Annunciation St
New Orleans, LA 70115-1848
919-599-7245
No fax available
william@livingcapitalmetrics.com

Key words

psychosocial metrology, measurement, reference standards, traceability, management metrics

CRITICAL NATIONAL NEED IDEA
METROLOGICAL INFRASTRUCTURE FOR HUMAN, SOCIAL, AND NATURAL CAPITAL

Introduction

A metrological infrastructure for human, social, and natural capital is an area of critical national need with potential for transforming education, health care, and the economy at large. The societal challenges preventing the development of reference standards and universally uniform metrics for these forms of capital are formidable. Societal challenges include

- the huge cost of developing these metrics;
- a general lack of awareness of the decades of research proving their viability and special advantages;
- an underdeveloped public appreciation for both the high returns provided by investments in metrology (NIST, 2003) and the vital role played by metrology in the history of science and capitalism (Ashworth, 2004);
- institutional orientations better able to serve the needs of existing paradigms than the emergence of new ones; and
- deeply rooted cultural presuppositions about the nature of number and the alleged limits of psychosocial measurement.

The economy, education, health care reform, and the environment are all now top priorities for attention. Interest in ecologically sustainable and socially responsible business practices has never been stronger or more widespread. Transparency and accountability are the key words of the day. Efforts are underway to tie health care reimbursements and executive compensation to outcome or impact measures. The assessment of risk, employee and organizational performance, environmental quality, and quality of life are all of increasing interest as every sector of the economy focuses on ways of preventing future debacles and on finding new sources of viable profits. Strong interest is being shown in new measures of overall economic performance interpreted less in terms of any and all commerce, no matter how detrimental to overall quality of life, and more in terms of genuine progress, happiness, and social cohesion (Anielski, 2007).

There is nothing more essential to succeeding in these efforts than the quality of the measures we develop and deploy. Even so, few, if any, of these efforts are taking advantage of longstanding, proven measurement technologies that may be crucial to the scientific and economic successes we seek. Bringing these technologies to the attention of the academic and business communities for use, further testing, and development in new directions is an area of critical national need.

The Technologies To Be Developed

Researchers in the history, philosophy, and social studies of science and technology have focused on metrology as a domain of central importance (Latour, 1987, 2005; Wise, 1995). There are two phases in the development of universally uniform metric systems. The first determines whether something exists in a persistent and stable measurable state independent of the sample measured, the equipment and operator measuring, time, and space. In this first phase, things themselves act as agents compelling agreement among observers as to their separate and real status as objective phenomena in the world.

The second phase in the process of establishing metrological uniformity transforms this agent of agreement into a product of agreement. Now, given a measurable phenomenon, research technicians collaborate on the unit size and range, nomenclature, and terminology by which its quantitative and qualitative features will be communicated. Systems for calibrating instruments in the standard metric, and checking their traceability to it, are devised and implemented.

The first phase in the process of establishing metrological uniformity for intangible forms of capital measured via ability tests, surveys, assessments, and ratings has effectively been underway for at least 50 years, since the work of Rasch (1960), and for more than 80 if Thurstone's (1928) pioneering work is included. Though virtually unknown outside psychometric circles, the facts of additive, independent, transitive, linear, ratio, and separable parameters for constructs measured in the human and social sciences are not controversial.

Much of this work has been conducted by researchers trained in the natural sciences who turned their attention to the social sciences, such as Thurstone (an electrical engineer), Rasch (a mathematician), and Wright (a physicist who worked as an assistant to Nobelists Townes and Mulliken before taking up psychometrics). A number of reviews of this work are available (Bezruzcko, 2005; Bond and Fox, 2007; Wright, 1997), with one appearing recently in an international journal of physical metrology (Fisher, 2009) along with another surveying the challenges involved (Finkelstein, 2009).

There are, however, few signs of any research programs, funding, or practical demands targeting the second phase in the work required for universally available metrologic uniformity in the measurement of intangible forms of capital. The viability of such a goal is suggested in a very small number of published research articles (Fisher, 1997a, 1997b, 2000, 2009a, 2009b; Smith & Taylor, 2004) but every calibration of an instrument producing data that meets demands for separable, independent model parameters—and more are published every day—provides support for another construct that is a possible candidate for a standardized metric.

A critical national need exists for a widespread awareness of the viability and desirability of reference standard metrics for human, social, and natural capital. Since the publication of a seminal paper in the early 1960s (Arrow, 1963), many economists have taken it for granted that health care is one industry in which common product definitions are impossible. The success of advanced measurement applications in health care research over the last 30 years (Bezruczko, 2005) contradicts that assumption.

We need to question the assumption that measurement in the social sciences is permanently relegated to an inferior status. We need research exploring possibilities for more rigorously defined metrics and the benefits that could be obtained from them. Many scientists in other fields, and many otherwise well-informed members of the public will be surprised to learn that the research that has quietly accrued over the last 50 years presents us with serious and significant potentials for rapid new productivity gains. The technologies to be developed should focus on:

- making the calibration of high quality instrumentation the norm across academia and business, wherever ability tests, attitude surveys, or performance assessments are used;
- educating researchers, business people, and the public as to the value of metrological markets trading in the common currencies of unified metrics;
- creating expectations that any measure worthy of the name is properly calibrated, expressed in a universal metric, with demonstrated and maintained traceability to it;
- thinking global while acting local, making individual measures, group averages, and aggregate indexes comparable, in the manner of the metric systems' nano-, milli-, centi-, and kilo-meters;
- enabling everyone to know where they stand relative to everyone else in terms of their available stocks of intangible capital resources, via personalized, neighborhood, community, state, regional, national, and international Genuine Progress or Happiness Indexes.
- fitting standardized metrics for all vital forms of intangible capital into econometric models, accounting frameworks, and financial spreadsheets;

- devising financial instruments, such as currencies, stocks, and bonds, for representing the amounts and value of the human, social, and natural capital possessed, owned, or invested in; and
- bringing all major stakeholders with interests in each major capital domain to the table, so that instruments, metrics, and applications can be tuned harmoniously across economic and social sectors and levels.

Expected New Outcomes and Capabilities

The primary results that could be obtained in an economic context informed by universally uniform, linear and ratio metrics for intangible forms of capital follow from the oft-repeated saying, "You manage what you measure." Most of the metrics currently used in the management of human, social, and natural capital are nonlinear, ordinal scores, ratings, and percentages. Because their unit magnitudes are dependent on locally variable score distributions, these alleged "metrics" are often uninformative, confusing, or deceptive (Murray, 2006; Ho, 2008). The incommensurability of these so-called measures effectively locks up human, social, and natural capital markets by making individual information transactions so expensive that decisions are made with no information, or with the wrong information.

The Table below contrasts the features of scores, ratings, and percentages with those of calibrated measures. Though advanced measurement applications have demonstrated these outcomes and capabilities for decades, they have not become the mainstream paradigm. And improved measurement is not an end in itself. Though we rarely stop to think about it, we all know that fair measures are essential to efficient markets. When different instruments measure in different units, market transactions are encumbered by the additional steps that must be taken to determine the value of what is being bought and sold. Health care and education are now so hobbled by myriad varieties of measures that common product definitions for the outcomes of these industries seem beyond reach.

As has been pointed out in a wide variety of works over the last several decades, we need to broaden the focus of business management beyond investments, factories, equipment, property, and labor. Instead of the traditional three forms of capital (land, labor, and manufactured equipment), we actually employ four (natural, human, manufactured equipment, and social) (Hawken, Lovins, and Lovins, 1999, p. 4; Ekins, Hillman, Hutchison, 1992, pp. 48-61). Land and labor are far more complex than a mere piece of ground and the functionality of a job description. These complexities are captured by the multifaceted concepts of natural and human capital, which have to include diverse and distinct dimensions of the resources brought to bear. And social capital is of such vital importance that capitalism itself could not have gotten off the ground without it.

In order to make capitalism live up to its own accounting principles, we need better measures fit into an accounting framework that redefines profit so that it is less a matter of liquidated capital, and more a matter of removing wasted resources from within a closed system of limited capacities. A book widely hailed for its description of the next industrial revolution suggests that waste is the common root cause of human suffering, sociopolitical discontent, and environmental degradation (Hawken, Lovins, and Lovins, 1999, p. 59). In order to learn how to reduce the waste of human potential, community trust, and natural resources, we must better learn the truth of the maxim, we manage what we measure.

Calibrated instruments traceable to reference standards express value in universally uniform metrics that function as common currencies. New efficiencies for human, social, and natural capital markets come from the reduced friction in transactions, which are made meaningful and comparable via metrological networks not much different from the one connecting all the clocks. The intangibles of health care, education, social services, and human/natural resource management may not be forever doomed to locally dependent product definitions that defy pricing.

What we need are ways of extending the basic capitalist ethos into the domain of the intangibles. How can we set up markets so that the invisible hand efficiently promotes social and environmental ends unintended by individuals maximizing their own gains? How might we extend the free play of self-interest into more comprehensively determined returns for the global dividend? Better measurement will inevitably be of central concern in answering these questions.

Imagine a world in which everyone everywhere has or can easily obtain comparable measures of their available literacy, numeracy, health, and community capital. The market value of each is readily available, alongside the Dow Jones, the S&P 500, and the daily changes in the price of gold and wheat. Students, parents, teachers, researchers, school accreditors, and investors track literacy and numeracy stock values over time and across schools, facilitating school choice and investment decisions. Patients, advocates, quality improvement specialists, purchasers, and researchers similarly track health capital stock values across healthcare providers, facilitating purchasing and treatment choice decisions.

In this context, the existence of an actual market of shared uniform information would coordinate the collective decisions of purchasers and providers to match supply and demand far more efficiently than could ever be the case in the current system of high-friction, ordinal, and locally dependent "metrics."

Perhaps an alternative tax structure would evolve, such that corporations are held to a minimum social capitalization. Instead of paying taxes to various levels of government to support schools, public health initiatives, fire and police protection, etc., corporations and investors would fund these directly by buying shares in the available literacy, numeracy, skill, motivation, and health stocks. Instead of today's method of hiring workers for pay and benefits, employers might instead invest in individual stocks of skills, motivations, and health under a variety of profit-sharing arrangements. Instead of having a job, workers would have their capital stocks on the market. And though there would be some incentive to shift loyalties to the highest bidder, the desire to increase the value and stock of social capital, and to maximize returns on investments in it, would provide considerable motivation toward stability.

Innovation is increasingly seen as best conceived as a group effort. The wisdom of crowds phenomenon makes it possible for actors coordinated by shared information to accomplish in short order tasks that either could not be done by independent individuals at all, or only by using much more time and resources. The profit motive is an energy source of incredible power and potential. Creating an economic and social context in which innovation on the broadest scale could be brought to bear on issues of human, organizational, and environmental performance and management would be productively disruptive and transformative on the highest levels.

Paths to Achieving Goals

Maps to Administration Guidance

For decades, economists, academics, and social and environmental activists have sought to bring the externalities of human, social, and natural capital under more direct management, and to have their values included in econometric models, accounting frameworks, policy formation, and decision making. The current Administration and Congress are making improved management of these forms of capital in education, health care, and environmental resource management a national priority. Congress is progressing with a cap-and-trade system of carbon credits, and large numbers of reports and research studies call for new and improved metrics, comparable effectiveness research in health care, and new infrastructural capacities (for one example, see Center for American Progress, 2009). Unfortunately, these efforts typically make unexamined and obsolete suppositions concerning the available measurement technologies, and so fail to aim at highly desirable goals and objectives that cannot be attained using outmoded technologies.

For instance, a consortium of foundations recently produced a report focusing on the conditions for entrepreneurial innovation in education (CAP, 2009). The report deplores the lack of a quality improvement culture in education and the general failure to recognize the vital importance of measuring performance for active management. The specific recommendations concerning power metrics and drastic improvements in information quality do not, however, recognize or build on existing capacities in the educational system. The report's recommendations would be quite different if the entrepreneurial focus had been blended with the technical capacities of educational measurement technologies that have been in place for decades.

The obvious recommendation with which to start concerns the reason why public education in the United States is such a fragmented system: because outcome standards and product definitions are expressed (almost) entirely in terms of locally-determined content and expert opinion. Local content, standards, and opinions are essential, but to be meaningful, comparable, practical, and scientific they have to be brought into a common scale of comparison.

The technology for creating such scales is widely available. For over 40 years, commercial testing agencies, state departments of education, school districts, licensure and certification boards, and academic researchers have been developing and implementing stable metrics that transcend the local particulars of specific tests. The authors of the "Stimulating Excellence" report are right to stress the central importance of comparable measures in creating an entrepreneurial environment in education, but they did not do enough to identify existing measurement capabilities and how they could help create that environment.

For instance, all three of the recommendations made at the bottom of page 12 and top of page 13 of the "Stimulating Excellence" report address capabilities that are already in place in various states and districts around the country. The examples that come easiest to mind involve the Lexile Framework for Reading and Writing, and the Quantile Framework for Mathematics, developed by MetaMetrics, Inc., of Durham, NC (www.lexile.com). A similar and extensive body of work has been accomplished by the Northwest Evaluation Association (NWEA; www.nwea.org).

The Lexile metric for reading ability and text readability unifies all major reading tests in a common scale, and is used to report measures for over 28 million students in all 50 states. Hundreds of publishers routinely obtain Lexile values for their texts, with over 115,000 books and 80 million articles (most available electronically) Lexiled to date.

Furthermore, though one would never know from reading the "Stimulating Excellence" report, materials on the MetaMetrics web site show that the report's three recommendations concerning the maximization of data utility have already been recognized and acted on, since

- many standardized assessments are already aligned with state learning standards,
- available products already quickly incorporate assessment results into the process of teaching and learning (and a lot more quickly than "a day or two after testing"!), and
- several states already have years of demonstrated commitment to keeping their standards and assessments relevant to the changing world's demands on students.

That said, a larger issue concerns the need to create standards that remain invariant across local specifics. A national curriculum and national testing standards will likely either dictate specific content or foster continued fragmentation when states refuse to accept that content. But in the same way that computer-adaptive testing creates a unique examination for each examinee—without compromising comparability—so, too, must we invest resources in devising a national system of educational standards that both takes advantage of existing technical capabilities and sets the stage for improved educational outcomes.

That is what the report's key recommendation ought to have been. An approximation of it comes on page

35, with the suggestion that now is the time for investment in what is referred to as “backbone platforms” like the Internet. Much more ought to have been said about this, and it should have been integrated with the previous recommendations, such as those concerning information quality and power metrics. For instance, on page 27, a recommendation is made to“build on the open-source concept.” Upon reading that, it seemed that the authors were going to make an analogy with adaptively administered item banks, not literally recommend actual software implementation processes.

But they took the literal road and missed the analogical boat. That is, we ought to build on the open-source concept by creating what might be called crowd-sourced wikitests—exams that teachers and researchers everywhere can add to and draw from, with the qualification that the items work in practice to measure what they are supposed to measure, according to agreed-upon data quality and construct validity standards. This process would integrate local content standards with global construct standards in a universally uniform metric not much different from the reference standard units of comparison we take for granted in measuring time, temperature, distance, electrical current, or weight.

And this is where the real value of the “backbone platform” concept comes in. The Internet, like phones and faxes before it, and like alphabetic, phonetic and grammatical standards before them, provides the structure of common reference standards essential to communication and commerce. What we are evolving toward is a new level of complexity in the way we create the common unities of meaning through which we achieve varying degrees of mutual understanding and community.

Measurement plays a fundamental role in the economy as the primary means of determining the relation of price to value. The never-ending spiral of increasing costs in education is surely deeply rooted in the lack of performance metrics and an improvement culture. We ought to take the global infrastructure of measurement standards as a model for what we need as a “backbone platform” in education.

We ought to take the metaphor of transparency and the need for “clear metrics” much more literally. We really do need instruments that we can look right through, that bring the thing we want to see into focus, without having to be primarily concerned with which particular instrument it is we are using. This is the implicit intention of many studies, reports, and policy recommendations that are made in the absence of specific information about and understanding of the available measurement technologies. Transparency is more than the mere availability of undigestible volumes of uninterpretable data; it has to provide a clear view on operations. Accountability has to be more than a responsibility for numbers that provide more of an appearance of objectivity than actual actionable value.

Justification for Government Attention

The societal challenges associated with the development and deployment of a metrological infrastructure for human, social, and natural capital are of a magnitude that prevent even the largest corporations or research institutes from undertaking the task alone. If the same historical process as occurred in the emergence of existing metrological standards were allowed to take its course, we could expect that the ongoing use of incommensurable metrics would eventually produce such discontent that a clamor for metric unification would arise of its own accord, as happened in the French Revolution.

Hope that such a repetition of history will occur is slight, given the decades that have passed with no emergent awareness that different tests, surveys, and assessments intended to measure the same things could in fact be designed to report out in the same units. Even growing awareness of the role of networks, common languages, and distributed cognition in innovation seems to be lost on those who could most benefit from a system of psychosocial metrology. Plainly, instead of letting this process happen to us in its own time, we need to make it happen in our own time.

If these challenges are not met, we will continue with an incomplete, unsustainable, and irresponsible economic system, one that effectively burns its resources and calls the results profits. Conservatively estimated, the liquid and manufactured capital, property, and labor costs that are included in current economic models and accounting spreadsheets comprise barely 10% of the actual total economy (Hawken, Lovins, and Lovins, 1999, pp. 5-6). That is, 90% of the capital resources circulating in the economy that make profits possible are not on the books. These include the irreplaceable air and water services provided by nature, and the absolutely essential trust, loyalty, and commitment provided by society.

As is increasingly widely recognized, it is impossible to continue with this system. But piecemeal efforts often are not just inadequate to the task, they actually make things worse as uncoordinated and mutually contradicting efforts compete for attention and resources. We need broad efforts undertaken by society as a whole, with everyone's interests represented. No individual, small business, major corporation, or nonprofit foundation could ever hope to succeed alone in a task of this scale and scope.

Likely proposers to a funding competition in this area would include commercial agencies already making use of the available advanced measurement techniques. In education, these include the Northwest Evaluation Association; MetaMetrics, Inc.; Educational Testing Service; ACT; Pearson; and many others. State departments of education with longstanding expertise in this area include those in Oregon, Michigan, Illinois, Vermont, and in many other states. In health care, research groups with relevant expertise include QualityMetric, FOTO, Inc., the Rehabilitation Institute of Chicago, PeaceHealth, and others.

Similarly, academic departments of psychology, education, sociology, public health, health systems management, and others conducting research and teaching in this area would also likely propose projects for funding. Academic departments with particularly high profiles in this area can be found at the University of California, Berkeley; University of Illinois, Chicago; Johns Hopkins University; Emory University; Northwestern University; Boston College; University of Denver; University of Michigan; the Chicago Medical School; and elsewhere.

Essentials for TIP Funding

The societal challenges related to the improved measurement of human, social, and natural capital are not being addressed for a number of reasons. First, despite the longstanding existence, of data, instruments, and theory to the contrary, it is widely and mistakenly believed that the fundamental measurement of constructs measured by way of ordinal observations is impossible. Second, it is also widely and mistakenly believed that all numbers are inherently and always effectively quantitative, or that supposed differences in kinds of numbers are purely academic and of no practical consequence. Third, the metrological infrastructure is almost completely invisible to and taken for granted by the public, meaning that local efforts aimed at expanding it into a new domain are virtually meaningless and inevitably futile. Fourth, the existing system of incentives and rewards makes it very difficult, if not impossible, for individual researchers and teachers to have an impact on the behaviors and decisions of their clients and students, as these are culturally rooted in the familiar, albeit misunderstood and misapplied, ordinal and local systems. Fifth, even when an individual or organization does grasp the significance of the new measurement technologies, these very few isolated and uncoordinated instances depend too heavily on local leadership, and eventually starve for lack of sustenance from a larger networked metrological culture.

The sum meaning of these five conditions is that research and development of metrological infrastructure for intangible capital will not proceed at all without leadership at the national level and without public funding. But the nation's scientific frontiers and commercial frontiers would potentially be greatly stimulated simply by publicly introducing the idea that such an infrastructure could be possible, and by suggesting that considerable benefits relative to existing goals for improving education, health care reform, and environmental management could accrue from it.

Scientists and technicians in many fields who work in academia or human or environmental resource management mistakenly assume that their own measures of educational achievement or organizational performance cannot live up to widely accepted scientific expectations as to instrument quality and the inferential stability of measures. Informing these scientifically sophisticated researchers, managers, and teachers of the new metrological horizons available to them could itself be transformative insofar as the seeds of a new metrological culture would be planted in the most fertile available ground with the most likely resources for ongoing nurturing and growth.

The concept of metrics for human, social, and natural capital traceable to reference standards will succeed or fail to the extent

- that all interested stakeholder groups are represented,
- that criteria for instrument calibration and the metrological repeatability and reproducibility studies are sufficient to the task,
- that applications in each domain (education, health care, human or environmental resource management, organizational performance and risk assessment, etc.) employ the same criteria, are implemented with full fidelity, and are coordinated relative to one another;
- that new metrological languages and cultures are built around the practical issues emerging in front line applications, such that the full meaning of the information obtained is employed, shared, recorded, and acted on at every level; and
- that the creation of more efficient markets for human, social, and natural capital harnesses the energy of the profit motive by revealing new opportunities for reducing waste, for improving returns on investment, and for entrepreneurial innovation.

This research could impact the nation in a transformational way by revealing the extent to which the apparent choice between capitalism and socialism is a false dichotomy. Civil infrastructure and social services need not be conceived as requiring direct social investment, with returns dispersed unevenly precisely because they are disseminated broadly to any and all who are in a position to benefit from them. Situated in a market system incorporating rigorously constructed and maintained common currencies for the exchange of measured value, the growth of human, social, and natural capital could be less a burden restricting profitability and more a direct source of profits. Instead of corporate welfare we could have public investment supporting the provision of needed infrastructure, with profits defined as much in terms of public benefit and the greater good as in terms of financial returns.

Even partial success could provide dramatic benefits. Imagine that the only metrics developed and deployed concern those essential to the health care industry. Further imagine that only a fraction of the returns on metrological improvement investments seen historically by NIST (2003) are obtained. Even at this very limited degree of success, a 10% annual return could mean that health care costs have stopped increasing, that health insurance coverage has become more affordable, and that the industry as a whole has become more profitable. Instead of health care costs doubling as a fraction of total GDP in the next 30 years, they could more likely be halved. And this would be accomplished in ways improving the quality of care involving the removal of waste, inconsistencies, and duplications, not as a result of unilaterally imposed lower payments. It would seem reasonable to hope for far more benefit to the nation, given the returns on investment historically obtained by NIST and given the introduction of an entirely new market context for entrepreneurial innovation. We won't, of course, find out what we're capable of until we try.

References

Anielski, M. (2007). *The economics of happiness: Building genuine wealth.* Gabriola, British Columbia: New Society Publishers.

Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. *American Economic Review*, *53*, 941-73.

Ashworth, W. J. (2004, 19 November). Metrology and the state: Science, revenue, and commerce. *Science, 306*(5700), 1314-7.

Bezruczko, N. (Ed.). (2005). *Rasch measurement in health sciences.* Maple Grove, MN: JAM Press.

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

Center for American Progress, The American Enterprise Institute, The Broad Foundation, New Profit Inc., PublicImpact, and the Annie E. Casey Foundation. (2009, May). *Stimulating excellence: Unleashing the power of innovation in education*. Retrieved 12 July 2009, from Center for American Progress: http://www.americanprogress.org/issues/2009/05/entrepreneurs_event.html.

Ekins, P., Hillman, M., & Hutchison, R. (1992). *The Gaia atlas of green economics (Foreword by Robert Heilbroner).* New York: Anchor Books.

Finkelstein, L. (2009). Widely-defined measurement—An analysis of challenges. *Measurement*, p. in press.

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, June). What scale-free measurement means to health outcomes research. *Physical Medicine & Rehabilitation State of the Art Reviews*, pp. 357-373.

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

Fisher, W. P., Jr. (2009a). Bringing human, social, and natural capital to life: Practical consequences and opportunities. In M. Wilson, K. Draney, N. Brown & B. Duckor (Eds.), *Advances in Rasch Measurement, Vol. Two* (p. in press). Maple Grove, MN: JAM Press.

Fisher, W. P., Jr. (2009b, July). Invariance and traceability for measures of human, social, and natural capital: Theory and application. *Measurement (Elsevier)*.

Hawken, P., Lovins, A., & Lovins, H. L. (1999). *Natural capitalism: Creating the next industrial revolution.* New York: Little, Brown, and Co.

Ho, A. D. (2008, August/September). The problem with "proficiency": Limitations of statistics and policy under No Child Left Behind. *Educational Researcher, 37*(6), 351-60.

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

Latour, B. (2005). *Reassembling the social: An introduction to Actor-Network-Theory.* Clarendon Lectures in Management Studies. Oxford, England: Oxford University Press.

Murray, C. (2006, Tuesday, July 25). By the numbers: Acid tests: No Child Left Behind is beyond uninformative. It is deceptive. *Wall Street Journal (New York),* http://www.opinionjournal.com/editorial/feature.html?id=110008701 [Last visited 11 July 2009].

NIST. (2003, 15 January). *Outputs and outcomes of NIST laboratory research*. Retrieved 24 April 2009, from http://www.nist.gov/director/planning/studies.htm#measures.

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.

Smith, R. M., & Taylor, P. (2004). Equating rehabilitation outcome scales: Developing common metrics. *Journal of Applied Measurement, 5*(3), 229-42.

Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology*, *XXXIII*, 529-544. (Reprinted in L. L. Thurstone, *The Measurement of Values*. Midway Reprint Series. Chicago, Illinois: University of Chicago Press, 1959, pp. 215-233.)

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. 352-61). Princeton, New Jersey: Princeton University Press.

Wright, B. D. (1997, Winter). A history of social science measurement. *Educational Measurement: Issues and Practice, 16*(4), 33-45, 52 [http://www.rasch.org/memo62.htm].

Table
Features of Scores, Ratings, and Percentages Compared with Measures

Features	Scores, Ratings, and/or Percentages	Calibrated Measures
Relation to sample distribution	Dependent	Independent
Paradigm	Descriptive statistics	Prescriptive measurement
Model-data relation	Models describe data, models fit to data, model with best statistics chosen	Models prescribe data quality needed for objective inference, data fit to models, GIGO principle
Relation to structure of natural laws	None	Identical
Statistical tests of quantitative hypothesis	None	Information-weighted and outlier-sensitive model fit, Principal Components Analysis, many other fit statistics available
Reliability coefficients	Cronbach's alpha, KR-20, etc.	Cronbach's alpha, KR-20, etc. and Separation, Strata
Reliability error source	Unexplained portion of variance	Mean square of individual errors
Range of measurement	Arbitrary, minimum-maximum score	Nonarbitrary, infinite
Unit status	Ordinal, nonlinear	Interval, linear
Unit status assumed in comparisons	Interval, linear	Interval, linear
Proofs of unit status	Correlational	Axiomatic; reproduced physical metrics; graphical plots; independent cross-sample recalibrations; etc.
Error theory for individual scores/measures	None	Derived from sampling theory
Architecture (capacity add/delete items)	Closed	Open
Supports adaptive administration and mass customization	No (changes to items change meaning of scores)	Yes (changes to items do not change meaning of measure)
Traceability to reference standard	No	Yes
Domains scored	Either persons or items but rarely both	All facets in model (persons, items, rating categories, judges, tasks, etc.)
Comparability of domains scored	Would be incomparable if scored	Comparable; each interpreted in terms of the other
Unscored domain characteristics	Assumed all same score or random	No unscored domain
Relation other measures same construct	Incommensurable	Commensurable and equatable
Construct definition	None	Consistency, meaningfulness, interpretability, and predictability of calibration/measure hierarchies
Focus of interpretation	Mean scores or percentages relative to demographics or experimental groups	Measures relative to calibrations and vice versa; measures relative to demographics or experimental groups
Relation to qualitative methods	Different philosophical commitments	Same philosophical commitments
Quality of research dialogue	Researchers' expertise elevated relative to research subjects	Subjects voice individual & collective perspectives on coherence of construct
Source of narrative theme	Researcher	Object of unfolding dialogue