urn:nbn:de:gbv:ilm1-2011imeko-018:1

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

A TECHNOLOGY ROADMAP
FOR INTANGIBLE ASSETS METROLOGY

William P. Fisher, Jr.[1], *A. Jackson Stenner*[2]

[1] Graduate School of Education, University of California – Berkeley, Berkeley, California, USA
[2]MetaMetrics, Inc., Durham, North Carolina, USA

Abstract – Measurement plays a vital role in the creation of markets, one that hinges on efficiencies gained via the universal availability of precise and accurate information on product quantity and quality. Fulfilling the potential of these ideals requires close attention to measurement and the role of technology in science and the economy. The practical value of a strong theory of instrument calibration and metrological traceability stems from the capacity to mediate relationships in ways that align, coordinate, and integrate different firms' expectations, investments, and capital budgeting decisions over the long term. Improvements in the measurement of reading ability exhibit patterns analogous to Moore's Law, which has guided expectations in the microprocessor industry for almost 50 years. The state of the art in reading measurement serves as a model for generalizing the mediating role of instruments in making markets for other forms of intangible assets. These remarks provide only a preliminary sketch of the kinds of information that are both available and needed for making more efficient markets for human, social, and natural capital. Nevertheless, these initial steps project new horizons in the arts and sciences of measuring and managing intangible assets.

Keywords measurement standards, technology roadmap, intangible assets

1. INTRODUCTION

Standards ensure the performance, conformity, and safety of innovative new products and processes. Manufacturing and the provision of services require standards to coordinate the matching of services (as in telecommunications), the fitting of parts, or the gauging of expectations [1]. Measurement, then, plays an essential economic role in the creation of markets centering on the efficiencies gained from the universal availability of precise, accurate, and uniformly interpretable information on product quantity and quality [2-5]. Clear, fully enforced property rights and transparent representations of ownership are other forms of standards that reduce the costs of transactions further by removing sources of unpredictable variation in social factors [6-9]. When objective measurement is available in the context of enforceable property rights and proof of ownership, economic transactions can be contracted most efficiently in the marketplace [10-11]. The emergence of objective measures of individual abilities, motivations, and health, along with service outcomes, organizational performance and environ-mental quality, present a wide array of new potential applications of this principle.

Proven technical capacities for systematic and continuous improvements in the quality of objective measures enable the alignment, coordination, and integration of expectations, investments, and capital budgeting decisions over the long term. The relationship between standards and innovation is complex and dynamic, but a general framework conducive to innovation requires close attention to standards. The trajectory of ongoing improvements in instrumentation in the psychosocial and environmental sciences suggests a basis for a technology road map capable of supporting the creation of new efficiencies in human, social, and natural capital markets. New efficiencies are demanded by macroeconomic models that redefine labour and land as human and natural capital, respectively, and that add a fourth form of capital—social—to the usual three-capitals (land, labour, and manufactured) framework.

These models enhance sensitivity to the full complexity of intangible assets, enable the conservation and growth of their irreplaceable value, and frame economics in terms of genuine progress, real wealth, sustainability, and social responsibility not captured in accounting and market indexes restricted to the value of property and manufactured capital. Of special interest is the fact that the technical features of improvements in rigorously defined and realized quantification are likely to be able to support the coordination of capital budgeting decisions in ways analogous to those found in, for instance, the microprocessor industry relative to Moore's Law.

The state of reading measurement [12-13] is sufficiently advanced for it to serve as a model in extrapolating the principle to further developments in the creation of literacy capital markets, and for generalizing the mediating role of instruments in creating markets to other constructs and forms of capital in the psychosocial, health, and environmental sciences.

Instruments, metrological standards, and associated conceptual images play vitally important mediating roles in economic success. For instance, the technology roadmap for the microprocessor industry, based in Moore's Law and its projection of doubled microprocessor speeds every two years, has success-

fully guided semiconductor market expectations and coordinated research investment decisions for over 40 years [2]. Moore's Law is more than a technical guideline—it has served as a business model for an entire industry for almost 50 years. This paper proposes the form similar laws and technology roadmaps will have to take to be capable of guiding innovation at both the technical level and at the broader level of human, social, and natural capital markets, comprehensively integrated economic models, accounting frameworks, and investment platforms.

The fulfilment of the potential presented by these intentions requires close attention to measurement and the role of technology in linking science and the economy [2-3]. Of particular concern is the capacity of certain kinds of instruments to mediate relationships in ways that align, coordinate, and integrate different firms' expectations, investments, and capital budgeting decisions over the long term.

Instruments capable of mediating relationships in these ways are an object of study in the social studies, history, and philosophy of science and technology. In this work, the usual sense of technology as a product of science is reversed [14-20]. Instead of seeing science as rigidly tied to data and rule-following behaviours, the term technoscience refers to a multifaceted domain of activities in which theory, data, and instruments each in turn serves to mediate the relation of the other two [21-23].

2. MEASUREMENT, MEDIATING INSTRUMENTS, AND MAKING MARKETS

In psychosocial research to date, there has been little recognition of the potential scientific and economic value of universally accessible, uniformly defined, and constant units. This article draws from the history of the microprocessor industry to project a model of how instruments measuring in such units can link science and the economy by coordinating capital budgeting decisions within and between firms. Links between the psychosocial sciences and industries such as education and health care are underdeveloped in large part because of insufficient attention to the mediating role some kinds of instruments are able to play in aligning investments across firms and agencies in an industry.

Instruments capable of mediating relationships do so by telling the story of a shared history and by envisioning future developments reliably enough to reduce the financial risks associated with the large investments required. In the microprocessor industry, for instance, Moore's Law describes a constant and predictable relation between increased functionality and reduced costs. From 1965 on, Moore's Law projected a detailed image of commercially viable applications and products that attracted investments across a wide swath of the economy. When it became clear in the early 1990s that the physical limits of existing technologies might disrupt or even end this improvement cycle, the Semiconductor Industry Association convened a special meeting aimed at creating a detailed common vision, a roadmap, for the next 15 years' developments in semiconductor technology [2].

This roadmap made it possible for the industry to navigate a paradigm shift in its basic technology with no associated economic upheaval and with the continuation of the historically established pattern of increased functionality and lower costs. Education, healthcare, government, and other industries requiring intensive human and social capital investments lack analogous ongoing improvements in their primary products' reliability, precision, and cost control. Where the microprocessor industry is able to reduce costs and improve quality while maintaining or improving profitability, education, healthcare, and social services seem only to always cost more, with little or no associated improvement in objective measures of quality.

To what extent might this be due to the fact that these industries have not yet produced mediating instruments like those available in other industries? If such instruments are necessary for articulating a shared history of past technical improvements and economies, and a shared vision of future ones, should not their development be a high priority? Within any economy, individual actors are able to contribute to the collective estimation of value only insofar as the information they have at hand is sufficient to the task. Ideally, with that information, those demanding higher quality can identify and pursue it, rewarding producers of the higher quality. Without that information, purchasers are unable to distinguish varying levels of quality consistently, so investments in improved products are not only unrewarded, they are discouraged. Philanthropic capital markets have lately been described in these terms [24].

Not yet having satisfactory mediating instruments in industries relying heavily on intangible assets is not proof of the impossibility of obtaining them. There are strong motivations for considering what appropriate mediating instruments would look like in human- and social-capital-intensive industries. Foremost among these motivations is a potential for correcting the significant capital misallocations caused when individual organizations make isolated investment decisions that cannot be coordinated across geographically distant groups' competing proprietary interests and temporally separated inputs and outputs.

The question is one of how to align investment decisions without compromising confidential budgeting processes or dictating choices. Simply sharing data on outcomes is a proven failure [25-26], and was never attractive to for-profit enterprises for which such information is of proprietary value. But instead of focusing on performance measured in locally idiosyncratic units incapable of supporting standard product definitions, might not a better alternative be found in defining a constant unit of increased learning, functionality, or health, and evaluating quality and cost relative to it? The key to creating coherent industry-wide com-

urn:nbn:de:gbv:ilm1-2011imeko-018:1

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

munities and markets is measurement. Fryback [27-28] succinctly put the point, observing that the U.S.

> health care industry is a $900 + billion [over $2.5 trillion in 2009 [29]] endeavor that does not know how to measure its main product: health. Without a good measure of output we cannot truly optimize efficiency across the many different demands on resources.

Quantification in health care is almost universally approached using methods inadequate to the task, resulting in ordinal and scale-dependent scores that cannot capitalize on the many advantages of invariant, individual-level measures [30]. Though data-based statistical studies informing policy have their place, virtually no effort or resources have been invested in developing individual-level instruments traceable to universally uniform metrics that define the outcome products of health care, education, and other industries heavily invested in human, social, and natural capital markets. It is well recognized that these metrics are key to efficiently harmonizing quality improvement, diagnostic, and purchasing decisions and behaviours [31]. Marshalling the resources needed to develop, implement them, and maintain them, however, seems oddly difficult to do until it is recognized that such a project must be conceived and brought to fruition on a collective level and against the grain of cultural presuppositions as to the objective measurability of intangible assets [32-33].

Probabilistic models used in scaling and equating different tests, surveys, and assessments to common additive metrics offer a body of unexamined resources relevant to the need for mediating instruments in the domains of human, social, and natural capital markets [33]. Miller and O'Leary [2] complement the accounting literature's overly narrow perspective on capital budgeting processes with the fruitful lines of inquiry opened up in the history, philosophy, and social studies of science. In this work, mathematical models and instruments are valued for their embodiment of the local and specific material practices through which mediation is realized.

In these practices, instruments capable of serving as reliable and meaningful media must simultaneously represent a phenomenon faithfully and facilitate predictable control over it. Though the philosophy of science has long focused attention on the nature of objective representation, the history and social studies of science have, over the last 30 years or so, shifted attention to the role of technology in theory development and in determining the outcome of experiments. By definition, instruments capable of mediating must exhibit properties of structural invariance across the locally defined contexts of different organizations' particular investments, policies, workforces, and articulations of the relevant issues. It is only through the conjoint processes of representation and intervention that, for instance, the steam engine became the medium facilitating development of work in the sense of engineering mechanics and in the economic sense of a new source of labour [34]. The medium is the message here, in the sense that mediating instruments like the steam engine both represent the lawful regularity of the scientific phenomenon and provide a predictable means of intervening in the production of it.

The unique importance and value of Rasch's models for measurement lie precisely here. Rasch-calibrated instruments have long been in use on a wide scale in applications that combine the representation of measured amounts for accountability purposes with instructional or therapeutic interventions that take advantage of the meaningful mapping of abilities relative to curricular or therapeutic challenges [35-38]. These models are structured as analogies of scientific laws' three-variable multiplicative form [13, 39] and so enable experimental tests of possible causal relations [40-41]. When data fit such a model, demonstrably linear units of measurement may be calibrated and maintained across instrument configurations or brands, and across measured samples. Clear thinking about the measured construct is facilitated by the invariant constancy of the unit of measurement—one more unit always means one more unit of the same size. When instruments measuring the same thing are tuned to the same scale, mediation is achieved in the comparability of processes and outcomes within and across subsamples of measured cases.

Linear performance measures are recommended as essential to outcome-based budgeting [11], and will require structurally invariant units capable of mediating comparisons in this way. Without instruments mediating meaningfully comparable relationships, it is impossible to effectively link science and the economy by coordinating capital budgeting decisions. The lessons so forcefully demonstrated over the course of the history of the microprocessor industry need to be learned and applied in many other industries.

The potential for a new class of mediating instruments resides here, where the autonomy of the actors and agencies forming a techno-economic network is respected and uncompromised. Rasch's parameter separation theorem is a scientific counterpart of Irving Fisher's economic separability theorem [42]. It is essential to realize that Rasch's equations model the stochastically invariant uniformity of behaviours, performances, or decisions of individuals (people, communities, firms, etc.), and are not statistical models of group-level relations and associations between variables [43]. Data fit a Rasch model and mediation is effected so far as the phenomenon measured (an ability, attitude, performance, etc.) retains its properties across samples and instrument brands or configurations. Given this fit, the unit of measurement becomes a common currency for the exchange of value within a market defined by the model parameters [42].

How could this implicit and virtual market be made explicit and actual? By devising mediating instruments linking separate actors and arenas in a way that conforms to the requirements of the techno-

urn:nbn:de:gbv:ilm1-2011imeko-018:1

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

economic forecasts of a projection like Moore's Law or of a technology roadmap based in such a law. As Miller and O'Leary [2] say,

> Markets are not spontaneously generated by the exchange activity of buyers and sellers. Rather, skilled actors produce institutional arrangements, the rules, roles and relationships that make market exchange possible. The institutions define the market, rather than the reverse.

What are the rules, roles and relationships that skilled actors need to arrange for their institutions to define efficient markets for human, social, and natural capital? What are the rules, the roles, and the relationships that make market exchange possible for these forms of intangible assets? How can standard product definitions for the outcomes of education, healthcare, and social services be agreed upon? Where are the lawful patterns of regularities that can be depended on to remain constant enough over time, space, firms, and individuals to support industry-wide standardizations of measures and products based on them? What trajectories can be mapped that would enable projections accurate enough for firms and agencies to rely on in planning products years in advance?

Answers to questions such as these provide an initial sketch of the kind of grounded, hands-on details of the information that must be obtained if the endless inflationary spirals of human- and social-capital-intensive industries are ever to be brought under control and transformed into profitable producers of authentic value and wealth.

3. THE RASCH READING LAW AND STENNER'S LAW

It is a basic fact of contemporary life that the technologies we employ every day are so complex that hardly anyone understands how they do what they do. Technological miracles are commonplace events, from transportation to entertainment, from health care to industry. And we usually suffer little in the way of adverse consequences from not knowing how automatic transmissions, thermometers, or digital video reproduction works. It is enough to know how to use the tool.

This passive acceptance of technical details beyond our ken extends as well into areas in which standards, methods, and products are much less well defined. And so managers, executives, researchers, teachers, clinicians, and others who need measurement but who are unaware of its technicalities tend to be passive consumers accepting the lowest common denominator of measurement quality.

And just as the mass market of measurement consumers is typically passive and uninformed, in complementary fashion the supply side is fragmented and contentious. There is little agreement among measurement experts as to which quantitative methods set the standard as the state of the art. Virtually any method can be justified in terms of some body of research and practice, so the confused consumer accepts whatever is easily available or is most likely to support a preconceived agenda.

It may be possible, however, to separate the measurement wheat from the chaff. For instance, measurement consumers may value a means of distinguishing among methods that emphasizes their interests in, and reasons for, measuring. Such a continuum of methods could be one that ranges from the least meaningful and generalizable to the most meaningful and generalizable, which is equivalent to ranging from the most to the least dependent on the local particulars of the specific questions asked, sample responding, judges rating, etc.

The aesthetics, simplicity, meaningfulness, rigor, and practical consequences of strong theoretical requirements for instrument calibration provide such criteria for choices as to models and methods [30. 44-49]. These criteria could be used to develop and guide explicit considerations of data quality, construct theory, instrument calibration, quantitative comparisons, measurement standard metrics, etc. along a continuum from the most passive and least objective to the most actively involved and most objective.

The passive approach to measurement typically starts from and prioritizes content validity. The questions asked on tests, surveys, and assessments are considered relevant primarily on the basis of the words they use and the concepts they appear to address. Evidence that the questions actually cohere together and measure the same thing is typically deemed of secondary importance, if it is recognized at all. If there is any awareness of the existence of axiomatically prescribed measurement requirements, these are not considered to be essential. That is, if failures of invariance are observed, they usually provoke a turn to less stringent data treatments instead of a push to remove or prevent them. Little or no measurement or construct theory is implemented, meaning that all results remain dependent on local samples of items and people. Passively approaching measurement in this way is then encumbered by the need for repeated data gathering and analysis, and by the local dependency of the results. Researchers working in this mode are akin to the woodcutters who say they are too busy cutting trees to sharpen their saws.

An alternative, active approach to measurement starts from and prioritizes construct validity and the satisfaction of the axiomatic measurement requirements. Failures of invariance provoke further questioning, and there is significant practical use of measurement and construct theory. Results are then independent of local samples, sometimes to the point that researchers and practical applications are not encumbered with usual test- or survey-based data gathering and analysis.

3.1. Six developmental stages
As is often the case, this black and white portrayal tells far from the whole story. There are multiple

urn:nbn:de:gbv:ilm1-2011imeko-018:1

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

shades of grey in the contrast between passive and active approaches to measurement. The actual range of implementations is much more diverse than the simple binary contrast would suggest. Spelling out the variation that exists could be helpful for making deliberate, conscious choices and decisions in measurement practice.

It is inevitable that we would start from the materials we have at hand, and that we would then move through a hierarchy of increasing efficiency and predictive control as understanding of any given variable grows. Previous considerations of the problem have offered different categorizations for the transformations characterizing development on this continuum. Stenner and Horabin [50] distinguish between 1) impressionistic and qualitative, nominal gradations found in the earliest conceptualizations of temperature, 2) local, data-based quantitative measures of temperature, and 3) generalized, universally uniform, theory-based quantitative measures of temperature.

The latter is prized for the way that thermodynamic theory enables the calibration of individual thermometers with no need for testing each one in empirical studies of its performance. Theory makes it possible to know in advance what the results of such tests would be with enough precision to greatly reduce the burden and expenses of instrument calibration.

Reflecting on the history of psychosocial measurement in this context, it then becomes apparent that these three stages can be further broken down. The distinguishing features for each of six stages in the evolution of measurement systems are expanded from a previously described five stage conception [12].

In Stage 1, conceptions of measurement are not critically developed, but stem from passively acquired examples. At this level, what you see is what you get, in the sense that item content defines measurement; advanced notions of additivity, invariance, etc. are not tested; the meanings of the scores and percentages that are treated as measures are locally dependent on the particular sample measured and items used; and there is no theory of the construct measured. Data must be gathered and analyzed to have results of any kind.

In Stage 2, measurement concepts are slightly less passively adopted. Additivity, invariance, etc. may be tested, but falsification of these hypotheses effectively derails the measurement effort in favour of statistical models with interaction effects, which are accepted as viable alternatives. Typically little or no attention is paid at this stage to the item hierarchy or the construct definition. An initial awareness of measurement theory is not complemented by any construct specification theory.

In Stage 3, measurement concepts are more actively and critically developed, but instruments still tend to be designed relative to content, not construct, specifications. Additivity and invariance principles are tested, and falsification of the additive hypothesis provokes questions as to why, where, and how those failures occurred. Models with interaction effects are

not accepted as viable alternatives, and significant attention will be paid to the item hierarchy and construct definition, but item calibrations remain empirical. Though there is more significant use of measurement theory, construct theory is underdeveloped, so no predictive power is available.

In Stage 4, the conceptualization of measurement becomes more active than passive. Initial efforts to (re-)design an instrument relative to construct specifications occur at this level. Additivity, invariance, etc. are explicitly tested and are built into construct manifestation expectations. The falsification of the additive hypothesis provokes questions as to why and corrective action, models with interaction effects are not accepted as viable alternatives, significant attention is paid to the item hierarchy and construct definition relative to instrument design, but empirical calibrations remain the norm. Some construct theory gives rise to limited predictive power. Commercial applications that are not instrument-dependent (as in computer adaptive implementations) exist at this level.

In Stage 5, all of the Stage 4 features appear in the context of a significantly active approach to measurement. The item hierarchy is translated into a construct theory, and a construct specification equation predicts item difficulties and person measures apart from empirical data. These features are used routinely in commercial applications.

In Stage 6, the most purely active approach to measurement, all of the Stage 4 and 5 features are brought to bear relative to construct specification equations that predict the mean difficulties of ensembles of items each embodying a particular combination of components. Commercial applications of this kind have been in development for several years.

Various degrees of theoretical investment at each stage can be further specified, along with speculations as to the extent of application frequency in mainstream and commercial instrument development. Stage 1, with no effective measurement or construct theory, remains the mainstream, most popular approach in terms of its application frequency, which likely exceeds 90 percent of all efforts aimed at quantifying human, social, or natural capital. It is, however, commercially the least popular in application frequency (<10 %?) in high stakes educational and psychological testing.

Stage 2, implementing very limited use of measurement theory and no construct theory is the next most popular mainstream psychosocial application frequency at perhaps eight percent, overall. It also has a somewhat higher commercial application frequency (10-20 %).

Stage 3, with a strong use of measurement theory and little or no construct theory, may be used as much as one or two percent of the time in mainstream applications, and may be dominant methodologically in commercial applications (55-65 %?).

Stage 4's strong use of measurement theory and use of some construct theory in informing instrument

urn:nbn:de:gbv:ilm1-2011imeko-018:1

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

design have very limited psychosocial application frequency in mainstream applications (<0,5 %?) but have made some significant starts in commercial applications (3-5 %?). Stage 5's strong theoretical understanding of constructs is virtually unknown in mainstream psychosocial application, but has also begun to see some commercial developments. Stage 6's mature theoretical understanding of constructs is only just emerging in some well-supported commercial applications.

3.2. The Rasch Reading Law

Measurement theory sets the stage for thinking about constructs by focusing attention on the meaningfulness of the quantities produced, by facilitating the construction of supporting evidence, by testing construct hunches, and by supporting theory development. Construct theory then sets the stage for following through on measurement theory's fundamental principles by making it possible to more fully transcend local particulars of respondent and item samples. It does so by recognizing that failures of invariance are valuable as anomalous exceptions that ―prove" (L. *probus*, test goodness of) the rule embodied in the measurement technology.

That is, data-model misfit is not considered to result from model failure, but from uninterpretable inconsistencies in the data stemming from underdeveloped theory and/or low quality data. Thus, failure to fit a model of fundamental measurement is not a sign of the end of the conversation or of the measurement effort. Rather, negative results of this kind provide needed checks on the strength of the object to withstand the rigors of propagation across media, which is the ultimate goal of having each different manufacturer's tool capable of functioning as a medium traceable to the same reference standard metric [18, 51].

The predictability of a trajectory for the evolution of measurement allows the specification of a law capable of shaping fundamental expectations as to increases in the power and complexity of psychosocial measurement technology, and the timing of those increases. This practical law is applicable to business relationships in a manner analogous to the way the basic law describes scientific relationships. This is so even if the definition of work in engineering mechanics is of little immediate interest in gauging the economic value of labour. Despite the lack of immediate relevance, the practical utility of the widely used horsepower measure of engine pulling capacity depends on the scientific validity of the proportionate relations between mass, force, and acceleration in Newton's laws.

The same simultaneous instantiation of scientific and economic value must be possible for instruments to mediate relationships in ways that can effectively and efficiently coordinate capital budgeting decisions. Thus, the Rasch Reading Law describes invariantly proportionate ratios between reading comprehension, text complexity, and reader ability [12-13]. As text complexity increases (the words used become less commonly encountered, and sentence length increases), reading comprehension rates decrease relative to a fixed reading ability measure. Conversely, given a fixed text complexity, reading comprehension rates increase as reading ability increases.

The practical value of this law is realized insofar as it then becomes possible to employ it productively in both (a) representing students' reading abilities in summative accountability measures and (b) intervening in ways likely to change those measures in formative instructional applications [35-38]. Concerning the latter, it is well understood that learning is inherently a matter of leveraging what is already known (the alphabet, numbers, words, grammar, arithmetical operations, etc.) to frame and understand what is not yet known (new vocabulary, constructions, specific problems, etc.). It is therefore vitally important to target instruction at the sweet spot where enough is known to support comprehension, but where what is not known is still substantial enough to make the lesson challenging. This range along the measurement continuum just above the student's measure is known as the Zone of Proximal Development [52] and is valued for indicating the range of curriculum content the student is developmentally ready to learn [53]. When measures are appropriately targeted, learning is maximized and measurement error is minimized. The same kind of strategy has proven useful in prescribing rehabilitation therapies [36] and likely has other as yet unexplored applications.

Targeting will be a key element in any future technology roadmap for education. Though there is no substitute for attention to other substantive aspects of the educational process, this indicator is of potentially central importance as a summary indicator of how accurately and precisely educational outcomes are represented, and how efficiently instructional interventions are implemented.

Rasch measurement isolates and focuses attention on empirical and theoretically tractable test item difficulty scale orders and positions. Then it estimates student abilities relative to that scale and describes them in terms of the probabilities of successful comprehension up and down the scale, *whether or not all of the items potentially available have actually been administered.* The goal of education, after all, is not to teach students only how to deal with the actual concrete problems encountered in instruction and assessment. The goal is rather to teach students how to manage any and all problems of a given type at a given level of difficulty.

Though a dialectic between part and whole is necessary, we cheat students and society when education becomes fixated on particular content and neglects the larger context in which skills are to be applied. The overall principle is effectively one of mass customization. Instruction and assessment, or any bidirectional
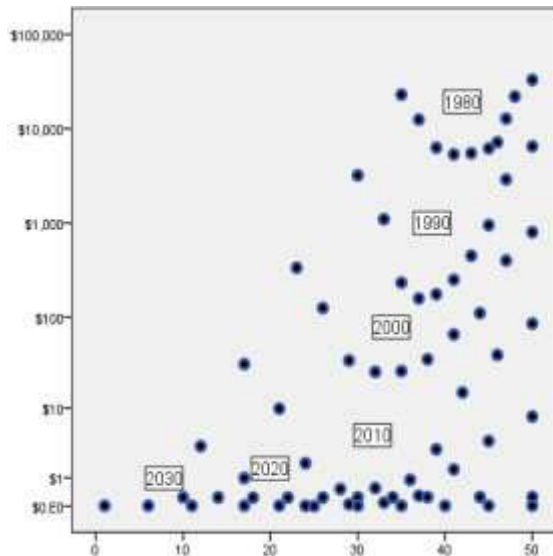
urn:nbn:de:gbv:ilm1-2011imeko-018:1

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

Fig. 1. Hypothetical Projection of Mean Targeting Accuracy (0-400L) By Average Relative US$ Cost of Producing a Single Precision Reading Measure By Year
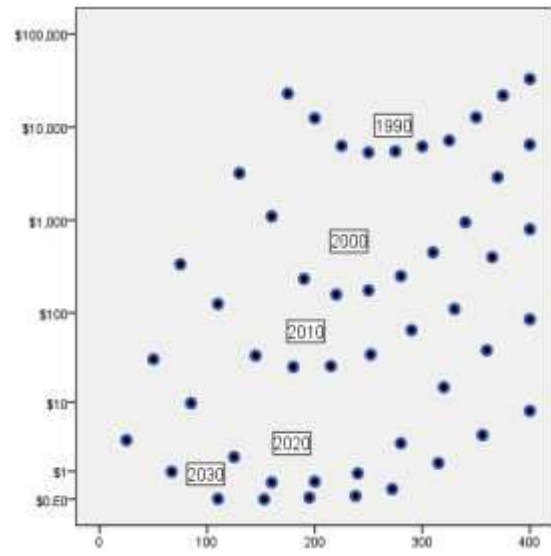

Fig. 2. Hypothetical Projection of Mean Percentages of Students Comprehending Text at a Rate of 60% or Less By Average Relative Cost of Producing a Single Precision Reading Measure By Year

method of simultaneous representation and intervention, benefits from forms of quantification coordinating substantive content with metrics that remain stable and constant no matter which particular test, survey, or assessment items are involved. The same principles apply in any other enterprise focused on intangible outcomes, such as health care, social services, or human resource management. We short change ourselves by failing to demand mediating instruments enabling a kind of virtual coordination of improvement, purchasing, hiring, and other investment decisions across different individuals, firms, agencies, and arenas in the economy. The architecture of probabilistic models open to the integration of new items and samples embodies the principles of invariance characteristic of the mediating instruments needed for aligning legally and geographically separated firms' decisions within a common inferential framework.

3.3. Stenner's law

Of course, even though it has been almost 60 years since Rasch [54] first did his foundational research [55-57] on reading, integrating assessment and instruction on the basis of the Rasch Reading Law is not yet the norm in educational practice. Accordingly, most instruction is not integrated with assessment, and few examination results are reported so as to illustrate the alignment of a developmental continuum with the curriculum. Furthermore, and more specifically, most reading instruction is not appropriately targeted at individual students' Zones of Proximal Development. This is problematic, given that reading abilities within elementary school classrooms can easily range from two grade levels below to two grade levels above the reading difficulty of the textbook.

Fig. 1, modelled on the first of two figures in Moore's original 1965 paper [58-59], shows a hypothetical but not unrealistic projection of the relation of average targeting accuracy with cost, by decade, from 1990 to 2030. Precision measurement is considered here to be realized when the targeted comprehension rate is realized to within 5%. Few reading tests were adaptively administered or well targeted before 1990; though a few were, computerization of test administration was difficult and expensive, as was (and remains) printed test production. Further, even fewer tests were administered for diagnostic or formative purposes before 1990, which is just as well as few would have been able to provide information useful for those applications.

It is plausible to suppose that, as the quality of testing has improved in the years after 1990, costs have been reduced and the targeting accuracy of assessment items and instructional text has been enhanced, so the difference between the average item difficulty and the average measure of the targeted student approaches 0, to the left. The upper limit of targeting accuracy remains constant because the impact of new methods of test construction and administration are unevenly distributed. Costs are driven down as theory is able to inform the automatic production and administration of targeted text and test items in computerized contexts effectively integrating assessment and instruction. Costs may be dropping by an order of magnitude every decade, with the rate in reductions in mistargeting slowing as it nears 0. At some future date, accurate targeting may become universal, and the right, off-target end of the range may also drop to near 0.

Fig. 2, also patterned on the first figure in Moore [58], is a variation on the same information as that shown in Fig. 1. Mistargeted text and test items may

urn:nbn:de:gbv:ilm1-2011imeko-018:1

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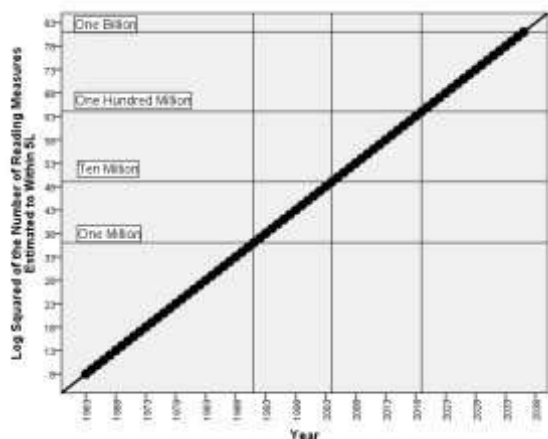urn:nbn:de:gbv:ilm1-2011imeko:2

Fig. 3. Rate of Increase in Number of Precision Reading Measures Estimated

bore able readers encountering material that is much too easy, but poor readers unable to make any headway with readings far too complex for them to comprehend are doomed to learn little or nothing. Fig. 2 is thus intended to convert Fig. 1's targeting information into the implied percentage of students comprehending text at a rate of 60% or less.

Fig. 3, patterned on the second figure in Moore [58], describes what may be referred to as Stenner's Law: the expectation that the number of precision reading measures estimated will double every two years, with no associated increase in cost. The figure has historical validity in that the line begins not long after the 1960 introduction of Rasch's work in Chicago, is in the range of 350,000 in the 1970s, during the Anchor Test Study [60-61], and is about 20-30 million in the period of 2005-2008, which is approximately how many measures were being produced annually at the time by users of the Lexile Framework for Reading [12].

4. A TECHNOLOGY ROADMAP FOR INTANGIBLE ASSETS

New and urgent demands challenged Moore's Law when it was realized in the 1990s that the physical limits of silicon could potentially disrupt the expectations that had allowed the microprocessor industry to coordinate its investment decisions so consistently for over 20 years [62]. The threat of a crisis led to the convening of an industry-wide meeting in 1992 by Gordon Moore, then chairman of the Semiconductor Industry Association's technology committee [2]. This and subsequent meetings of the group resulted in the annual publication of an International Technology Roadmap for Semiconductors.

These charts provided a level of specificity and detail not present in the more bare-bones projections of Moore's Law. The established history of past successes combined with new uncertainties compelled leaders in the field to seek out a basis on which new mediating instruments might be founded. Risks associated with evaluating several different methods of

resolving the technical problem of continued reductions in microprocessor size and cost had to be mitigated so that no firms found themselves making large capital investments with no product or customers in sight (which, unfortunately, is the status quo in education, healthcare, and other industries making intensive investments in human, social, and natural capital).

Table 1 presents a reading measurement variation on the 2001 version of the semiconductor industry's roadmap [2]. The basic structure of the table (the columns labelled ―Year of first production" and ―Technology node", and the subheadings focusing on ―Expected shifts in product functionality and cost", Introductory volumes, and Innovations) is identical with the one produced by the semiconductor industry. The remaining elements have been changed to focus on the kinds of functionality, cost, and innovations that have taken place historically in the domain of reading instruction and assessment.

Some of these suggested elements may prove less important in articulating a shared history and projecting an accessible vision of the future, and others may be needed. The point here is less one of specifying what exactly should be tracked and is more focused on conveying the general conceptual framework in which new possibilities for coordinating investment decisions in education might be explored. Plainly, it would be essential for the major stake holding actors and agencies involved in education, from academia to business to government, to themselves determine the actual contents of a roadmap such as this.

5. CONCLUSION

The mediation of individual and organizational levels of analysis, and of the organizational and inter-organizational levels, is facilitated by Rasch measurement. Miller and O'Leary [2] document the use of Moore's Law in the microprocessor industry in the creation of technology roadmaps that lay out the structure, processes, and outcomes that have to be aligned at all three levels to coordinate an entire industry's economic success. Such roadmaps need to be created for each major form of human, social, and natural capital, with the associated alignments and coordinations put in play at all levels of every firm, industry, and government.

It has been suggested that economic recovery in the wake of the Great Recession could be driven by a new major technological breakthrough, one of the size and scope of the IT revolution of the 1990s. This would be a kind of Manhattan Project or international public works program, providing the unifying sense of a mission aimed at restoring and fulfilling the promises of democracy, justice, freedom, and prosperity. Industry-wide systems of metrological reference standards for human, social, and natural capital fit the bill. Such systems would be a new technological breakthrough on the scale of the initial IT revolution. They would also be a natural outgrowth of existing IT sys-

urn:nbn:de:gbv:ilm1-2011imeko-018:1

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

Table 1. Sketch of a Possible Technology Roadmap for Literacy Education (Elements to be determined)

Technology node	Year of first production					
	1980	1990	2000	2010	2020	2030
Expected shifts in product functionality and cost—high volume targeted items						
Number of steps in item production and administration						
Multiple per 3-year technology cycle						
Affordable production cost per item at introduction						
Rate of cost reduction per cycle (%)						
Annualized cost reduction/item (%)						
Improvements in instructional targeting accuracy						
Reduction in students reading at <60% comprehension (%)						
Increase in number of precision reading measures (%)						
Available item volume at introduction						
Uncalibrated						
Calibrated bank						
From specification equation						
Innovations in item administration						
Paper and pencil adaptive formats						
Computer adaptive (bank)						
Theory adaptive (specified)						
Innovations integrating assessment & instruction						
Examinations aligned with curricula (%)						
Differentiated instruction aligned with adaptive tests (%)						
Writing assessment included (%)						

tems, an extension of existing global trade standards, and would require large investments from major corporations and governments. In addition, stepping beyond those suggestions that have appeared in the popular press, systematic and objective methods of measuring intangible assets would help meet the widely recognized need for socially responsible and sustainable business practices.

Better measurement will play a vital role in reducing transaction costs, making human, social, and natural capital markets more efficient by facilitating the coordination of autonomous budgeting decisions. It will also be essential to fostering new forms of innovation, as the shared standards and common product definitions made possible by advanced measurement systems enable people to think and act together collectively in common languages.

Striking advances have been made in measurement practice in recent years. Many still assume that assigning numbers to observations suffices as measurement, and that there have been no developments worthy of note in measurement theory or practice for decades. Nothing could be further from the truth.

Theory makes it possible to know in advance what the results of empirical calibration tests would be with enough precision to greatly reduce the burden and expenses associated with maintaining a unit of measurement. There likely would be no electrical industry at all if the properties of every centimetre of cable and every appliance had to be experimentally tested. This principle has been employed in measuring human, social, and natural capital for some time, but has not yet been adopted on a wide scale.

This might change with the introduction of Stenner's Law, and in conjunction with technology roadmaps for literacy capital and for other forms of human, social, and natural capital that project rates of increase in psychosocial measurement functionality and frame an investment appraisal process ensuring the ongoing creation of markets for advanced calibration services for the next 10 to 20 years or more.

A progression of increasing complexity, meaning, efficiency, and utility can be used as a basis for a technology roadmap that will enable the coordination and alignment of various services and products in the domain of intangible assets. A map to the theory and practice of calibrating instruments for the measurement of intangible forms of capital is needed to provide guidance in quantifying constructs such as literacy, health, and environmental quality. We manage what we measure, so when we begin measuring well what we want to manage well, we'll all be better off.

ACKNOWLEDGEMENT

REFERENCES

[1] R.H. Allen, R.D. Sriram, ―The role of standards in innovation", *Technological Forecasting and Social Change*, vol. 64, pp. 171-181, 2000.

[2] P. Miller, T. O'Leary, ―Mediating instruments and making markets: Capital budgeting, science and the

urn:nbn:de:gbv:ilm1-2011imeko-018:1

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

economy", *Accounting, Organizations, and Society*, vol. 32, no. 7-8, pp. 701-34, 2007.

[3] M. Callon, "From science as an economic activity to socioeconomics of scientific research: The dynamics of emergent and consolidated techno-economic networks", In P. Mirowski & E.-M. Sent (Eds.), *Science bought and sold: Essays in the economics of science.* Chicago: University of Chicago Press, 2002.

[4] Y. Barzel, "Measurement costs and the organization of markets", *Journal of Law and Economics*, vol. 25, pp. 27-48, 1982.

[5] A. Benham, L. Benham, "Measuring the costs of exchange", In C. Ménard (Ed.), *Institutions, contracts and organizations: Perspectives from new institutional economics* (pp. 367-375). Cheltenham, UK, Edward Elgar, 2000.

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

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

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

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

[10] G.P. Baker, R. Gibbons, K.J. Murphy, "Bringing the market inside the firm?", *American Economic Review* vol. 91, no. 2, pp.: 212-18, 2001.

[11] M.C. Jensen, "Paying people to lie: the truth about the budgeting process", *European Financial Management,* vol. 9, no. 3, pp. 379-406, 2003.

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

[13] D.S. Burdick, M.H. Stone, A.J. Stenner, "The Combined Gas Law and a Rasch Reading Law", *Rasch Measurement Transactions*", vol. 20, no. 2, pp. 1059-60, 2006.

[14] R. Bud, S. E. Cozzens (Eds.), *Invisible connections: Instruments, institutions, and science*. Bellingham, Washington, USA, SPIE Optical Engineering Press, 1992.

[15] 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.

[16] Y. M. Rabkin, Rediscovering the instrument: Research, industry, and education. In R. Bud & S. E. Cozzens (Eds.), *Invisible connections: Instruments, institutions, and science* (pp. 57-82). Bellingham, Washington, SPIE Optical Engineering Press, 1992.

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

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

[19] T.L. Hankins, R.J. Silverman, *Instruments and the imagination*. Princeton, New Jersey, Princeton University Press, 1999.

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

[21] J.R. Ackermann, *Data, instruments, and theory: a dialectical approach to understanding science*, Princeton, New Jersey, Princeton University Press, 1985.

[22] D. Ihde, *Instrumental realism: the interface between philosophy of science and philosophy of technology*, Bloomington, Indiana, Indiana University Press, 1991.

[23] D. Ihde, *Expanding hermeneutics: visualism in science*, Evanston, Illinois, Northwestern University Press, 1998.

[24] S.H. Goldberg, *Billions of drops in millions of buckets: Why philanthropy doesn't advance social progress.* New York, Wiley, 2009.

[25] C. Murray, "By the numbers: Acid tests: No child left behind is beyond uninformative. It is deceptive", *Wall Street Journal*, Tuesday, July 25, 2006.

[26] A.D. Ho, "The problem with 'proficiency': Limitations of statistics and policy under No Child Left Behind", *Educational Researcher*, vol. 37, no. 6, pp. 351-60, 2008.

[27] D. Fryback, "QALYs, HYEs, and the loss of innocence", *Medical Decision Making,* vol. 13, no. 4, pp. 271-2, 1993.

[28] D.A. Kindig, "Purchasing population health: Aligning financial incentives to improve health outcomes", *Nursing Outlook*, vol. 47, no. 1, pp. 15-22, 1999.

[29] "National Health Expenditure Data: NHE Fact Sheet." 2011, https://www.cms.gov/NationalHealthExpendData/25_NHE_Fact_Sheet.asp, Baltimore, MD, US, (Accessed 30 June 2011).

[30] D. Andrich, "Controversy and the Rasch model: A characteristic of incompatible paradigms?", *Medical Care*, vol. 42, no. 1, pp. I-7--I-16, 2004.

[31] D.M. Berwick, B. James, M.J. Coye, "Connections between quality measurement and improvement", *Medical Care*, vol. 41, no. 1 (Suppl), pp. I30-38, 2003.

[32] R.D. Cooter, "Law from order: Economic development and the jurisprudence of social norms", In *A Not-So-Dismal Science: A Broader View of Economies and Societies*, Eds. M. Olson and S. Kahkonen, pp. 228-44. New York, Oxford University Press, 2000.

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

[34] M.N. Wise, "Mediating machines", *Science in Context*, vol. 2, no. 1, pp. 77-113, 1988.

[35] A.C. Alonzo, J.T. Steedle, "Developing and assessing a force and motion learning progression", *Science Education*, vol. 93, pp. 389-421, 2009.

[36] W.C. Chang, C. Chan, "Rasch analysis for outcomes measures: some methodological considerations", *Archives of Physical Medicine and Rehabilitation*, vol. 76, no. 10, pp. 934-39, 1995.

[37] C.A. Kennedy, M. Wilson, "Using Progress Variables to Interpret Student Achievement and Progress.

urn:nbn:de:gbv:ilm1-2011imeko-018:1

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

Berkeley, Evaluation & Assessment Research Center. Paper Series, No. 2006-12-01. Berkeley: University of California, Berkeley BEAR Center. 51 pp., 2007.

[38] D. Leclercq, "Computerised tailored testing: Structured and calibrated item banks for summative and formative evaluation", *European Journal of Education*, vol. 15, no. 3, pp. 251-60, 1980.

[39] W.P. Fisher, Jr., "The standard model in the history of the natural sciences, econometrics, and the social sciences", *Journal of Physics, Conference Series,* vol. 238, no. 1, http://iopscience.iop.org/1742-6596/238/1/012016/pdf/1742-6596_238_1_012016.pdf, 2010.

[40] C.V. Bunderson, V.A. Newby, "The relationships among design experiments, invariant measurement scales, and domain theories", *Journal of Applied Measurement*, vol. 10, no. 2, pp. 117-37, 2009.

[41] A.J. Stenner, M. Smith III, "Testing construct theories", *Perceptual and Motor Skills*, vol. 55, pp. 415-26, 1982.

[42] W.P. Fisher, Jr., "Bringing human, social, and natural capital to life: Practical consequences and opportunities", *Journal of Applied Measurement*, vol. 12 no. 1, pp. 49-66, 2011.

[43] W.P. Fisher, Jr., "Statistics and measurement: Clarifying the differences", *Rasch Measurement Transactions*, vol. 23, no. 4, pp. 1229-1230, 2010.

[44] D. Andrich, "Understanding resistance to the data-model relationship in Rasch's paradigm: A reflection for the next generation", *Journal of Applied Measurement*, vol. 3, no. 3, pp. 325-59, 2002.

[45] I.J. Myung, "Importance of complexity in model selection", *Journal of Mathematical Psychology*, vol. 44, no. 1, pp. 190-204, 2000.

[46] I.J. Myung, M. Pitt, "Model comparison methods", In L. Brand & M. L. Johnson (Eds.), *Methods in Enzymology, Evaluation, Testing and Selection. Vol. 383: Numerical computer methods, part D* (pp. 351-65). New York, Academic Press, 2004.

[47] J.R. Busemeyer, Y.M. Wang, "Model comparisons and model selections based on generalization criterion methodology", *Journal of Mathematical Psychology*, vol. 44, no. 1, pp. 171-189, 2000.

[48] B.D. Wright, "A history of social science measurement", *Educational Measurement: Issues and Practice*, vol. 16, no. 4, pp. 33-45, 52, 1997.

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

[50] A.J. Stenner, I. Horabin, "Three stages of construct definition", *Rasch Measurement Transactions*, vol. 6, no. 3, p. 229, 1992.

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

[52] L.S. Vygotsky, *Mind and Society: The Development of Higher Mental Processes*. Cambridge, Massachusetts, Harvard University Press, 1978.

[53] P. Griffin, "The comfort of competence and the uncertainty of assessment", *Studies in Educational Evaluation*, vol. 33, pp. 87-99, 2007.

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

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

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

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

[58] G. Moore, "Cramming more components onto integrated circuits", *Electronics*, vol. 38, no. 8, pp. 114-117, 1965.

[59] G. Moore, "Progress in digital integrated electronics", International Electronics Devices Meeting (IEEE) *Technical Digest*, vol. 21, pp. 11-13, 1975.

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

[61] R.R. Rentz, E.L. Bashaw, "The National Reference Scale for Reading: An application of the Rasch model", *Journal of Educational Measurement*, vol. 14, no. 2, pp. 161-179, 1977.

[62] M. Schulz, "The end of the road for silicon?", *Nature*, vol. 399, pp. 729-730, 1999.

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

A. Jackson Stenner, MetaMetrics, Inc. (27713, Durham, North Carolina, USA) +1-919-547-3402, jstenner@lexile.com.