Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 114 (2017) 165–174

Procedia Computer Science

www.elsevier.com/locate/procedia

Complex Adaptive Systems Conference with Theme: Engineering Cyber Physical Systems, CAS October 30 – November 1, 2017, Chicago, Illinois, USA

A Practical Approach to Modeling Complex Adaptive Flows in Psychology and Social Science

William P. Fisher, Jr.*

*LivingCapitalMetrics.com, Sausalito, CA, USA; and BEAR Ctr, UC Berkeley, Berkeley, CA, USA

Abstract

Five moments in the formation and functioning of complex adaptive systems are: (1) emergent regularities and patterns in the flow of matter, energy, and/or information; (2) condensed schematic representations of these regularities enabling their identification; (3) reproductively interchangeable variants of these representations serving as templates for new instances of the pattern; (4) successful reproduction facilitated by the accuracy and reliability of the representations' predictions of data flow regularities; and (5) informational feedback that adaptively modifies and reorganizes representations to incorporate new variations in the data flow, cycling back the first moment. These five moments are instantiated via stochastic models providing practical approaches to representing and managing complex adaptive psychological and social systems in education, health care, human resource management, etc. Local independence, unidimensionality, and statistical sufficiency criteria function as means of identifying, evaluating, and deploying conceptual and social forms of life acting as evolving agents in defined ecological niches. Bringing these agents into play systematically requires embodying them in technologies instrumental to making them readily recognizable and sharable across ecosystem niches. Modeling research and practice promoting sustainable and self-organizing ecosystems of this kind set the stage for redefining profit in terms of authentic wealth and value for life.

Peer-review under responsibility of the scientific committee of the Complex Adaptive Systems Conference with Theme: Engineering Cyber Physical Systems.

Keywords: complex adaptive systems; psychology; stochastic models; Rasch models; self-organizing forms of life;

* Corresponding author. Tel.: +1-919-599-7245; .
E-mail address: wfisher@berkeley.edu

1877-0509 © 2017 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of the scientific committee of the Complex Adaptive Systems Conference with Theme: Engineering Cyber Physical Systems.
10.1016/j.procs.2017.09.027

1. Introduction

Living systems, from bacteria and viruses to plants, animals, and ecosystems to languages, societies, and economies, are complex adaptive systems characterized by five moments in their self-organizing processes [1]. Locating these five moments within a domain of practice, such as educational assessment or survey research, begins by tracing their embodiment through existing and proven research practices applying stochastic models of individual-level measurement [2-6]. Following the lead of recent research in agent-based modeling, this complex adaptive schema sets up a new practical art and science of instrument calibration and measurement that works from a bottom up flow of information. Instead of imposing policy from the top down on the basis of centralized data analyses, research and practice in a wide range of fields could perhaps more productively be provided media for their self-organization from the ground up.

2. Stochastic models and the opportunities they present

A broad array of stochastic models of individual-level measurement [2-6] present new opportunities for practical applications of agent-based models of complex adaptive systems [7-8] in designing, testing, and implementing information infrastructure ecosystems [9]. The local independence criterion tested by stochastic measurement models functions in effect as a way of identifying, evaluating, and deploying conceptual and social forms of life acting as agents in defined ecological niches. Bringing these agents into play systematically requires embodying them in technologies instrumental to making them readily recognizable, sharable, and habitually accessible. Such technologies could include common languages, shared metrics, and assessment instruments calibrated to be traceable to standard units of measurement, as shown in recent collaborations of metrology engineers and psychometricians [10-14].

Even when the methodological possibilities presented by adaptive instrument administration [15] and theory-informed unit standards are appreciated [16], the challenges in education, for instance, encountered in efforts aimed at bringing the world of varying curricula, pedagogies, assessments, and student abilities into a common framework of this kind may make such a framework seem impossibly unworkable. Decades of practical application of Rasch's models, however, suggest that coherent frames of reference aligning within-student development comparisons across students are feasible, viable, and desirable [17-21].

The phenomenon of noise-induced order [22], sometimes referred to as stochastic resonance [23], is a key characteristic of complex systems. Galison [24, pp. 843-844] encounters this kind of complexity in his extensive study of the material culture of microphysics, remarking on how the effectiveness of science stems from a kind of systematic disunity. By analogy, Galison points out that engineers have learned the value of amorphous semiconductors and laminated materials that fail microscopically but hold macroscopically when more homogenous materials collapse. The cultural import of this work stems from the fact that the disunity of science's communities of theoreticians, experimentalists, and instrument makers is not a function restricted or located solely in the domain of specific fields. It is, rather, fundamentally human in its incorporation of analogous linguistic and social complexities. Recognizing the pervasiveness of these patterns may provide a context for innovation that education and other fields could do well to emulate [25]. More fundamentally, the stochastic conception of error and uncertainty provides a basis for modeling meaningful quantitative interval units of measurement for emergent self-organizing forms of life [26].

2.1. Five moments in complex adaptive system emergence and function

The conception, gestation, birthing, and nurturing of complex adaptive systems constitute a reproductive logic for sociocultural traditions. Scientific traditions, in particular, form mature self-identities via a mutually implied subject-object relation absorbed into the flow of a mathematical dialectic. Complex adaptive systems establish the reproductive viability of their offspring and the coherence of an ecological web of meaningful relationships by means of this dialectic. Taylor [1, pp. 166-168] describes the five moments in the formation and operation of complex adaptive systems, which must be able:

- to identify regularities and patterns in the flow of matter, energy, and information (MEI, or data) in the environment;

- to produce condensed schematic representations of these regularities so they can be identified as the same if they are repeated;
- to form reproductively interchangeable variants of these representations;
- to succeed reproductively by means of the accuracy and reliability of the representations' predictions of regularities in the MEI data flow; and
- adaptively modify and reorganize representations by means of informational feedback from the environment.

All living systems, from bacteria and viruses to plants and animals to languages and cultures, are complex adaptive systems characterized by these five features.

In the history of science, technologically-embodied measurement functions as a complex adaptive system extending natural language's support for model-based reasoning and distributed collective cognition [27-28]. Each of Taylor's five moments in the formation and operation of complex adaptive systems describes a capacity of measurement systems, in that:

- data flow regularities are captured in initial, provisional instrument calibrations;
- condensed local schematic representations are formed when an instrument's calibrations are anchored at repeatedly observed, invariant values;
- interchangeable nonlocal versions of these invariances are created by means of instrument equating, item banking, metrological networks, and selective, tailored, adaptive instrument administration;
- measures read off inaccurate and unreliable instruments will not support successful reproduction of the data flow regularity, but accurate and reliable instruments calibrated in a shared common unit provide a reference standard metric that enhances communication and reproduces the common voice and shared identity of the research community; and
- consistently inconsistent anomalous observations provide feedback suggesting new possibilities for as yet unrecognized data flow regularities that might be captured in new calibrations.

Measurement in the social sciences is in the process of extending this functionality into practical applications in business, education, health care, government, and elsewhere. Over the course of the last 50 years, Rasch's probabilistic models for measurement research and practice have already iterated many times through these five moments.

2.2. Complex adaptive data flow regularities, schema, and compression

How? What does a "data flow regularity" look like? How is it condensed into a schematic and used to calibrate an instrument? How are local schema combined together in a pattern used to recognize new instances of themselves? More specifically, how might enterprise resource planning software (such as SAP, Oracle, or PeopleSoft) simultaneously provide both the structure needed to support meaningful comparisons and the flexibility needed for good fit with the dynamic complexity of adaptive and generative self-organizing systems? Prior work in this area proposes a dual-core, loosely coupled organization using ERP software to build social and intellectual capital, instead of using it as an IT solution addressing organizational inefficiencies [29]. The adaptive and generative functionality provided by probabilistic measurement models [2-5,15-16] makes it possible to model intra- and inter-organizational interoperability [30] at the same time it augments social and intellectual capital resources by bringing them to life socially and economically [31].

Fig. 1 (a) shows a sample data matrix of 24 columns by 40 rows. The individual observations are responses to questions (the columns) on a survey given by participants in a research study (the rows). These data are a small subset extracted randomly (without replacement) from a much larger data set involving hundreds of items and thousands of survey respondents. The larger data set was analyzed using the Winsteps Rasch measurement software [32] to calibrate an instrument measuring the quality of services in special education [33].

The individual responses shown in Fig. 1 are ordered so that the person with the highest measure and the highest probability of responding in the highest category (3) across all of the items is at the top. The person with the lowest measure and the lowest probability of responding in the highest category on any item is at the bottom. The survey item that is easiest to agree with and that garners the most responses of 3 ratings, is in the far left column, and the one that is hardest to agree with is on the far right.

To illustrate the consistency of the stochastic pattern in the data, responses of 3 are colored red, responses of 2 are black, and responses of 1 are blue. This pattern may not look like much to those unfamiliar with data of this kind

Fig. 1. Three samples of respondents (rows) answering the same questions (columns): (a) first subset; (b) second subset; (c) third subset.

evaluated in the context of probabilistic measurement theory, its capacity to take missing data into account, and the role of error ranges and confidence intervals in interpreting measures. It may seem as though this pattern is an accident that occurred once, never to appear again. As Taylor warns, the system has to be able to avoid the errors of mistaking order for randomness, and vice versa.

And so the question arises: to what extent is this pattern repeated across other samples from the larger data set? Fig. 1 (a), (b), and (c) show the same items (columns) in the same order, but with entirely different random samples of respondents (rows), as can be seen from the different entry order sequence numbers at the far left in each figure.

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

The same thing happens when the rows are held constant, and the columns are varied. In other words, the constancy of the pattern observed when different people answer the same questions is also found when the same people answer different questions. And so, as shown in Fig. 2, it also happens that the same patterns are also reproduced across

```
  99 + 333 33333 3333 33 3 3 3
 923 + 333 33333 3333 23 333 3
 915 +33  33   3333  33333  3
1110 + 333 33333 3332 32 3 3 3
1742 +33  32   1333  23333  3
1544 +33  33   3333   3333  1
2257 +33  33   2333  33332  1
 118 + 333 33331 2333 21 1 3 1
 753 +33  33   2323   12 3  1
1104 +33  22   2332  12222  2
2480 +32  32   2222  22322  3
 108 + 333 22322 2212 12 212 2
1269 +23  32   2322  22222  1
 114 +22  22   2222  22222  1
2587 +32   2   1122  21222
2242 +22  21   2221  12122  1
1590 + 22  322 1  221 11   1
1205 +32  22   2111  11111  1
 845 + 211 22111 1111 11 111 1
1917 + 111 11211 1111 11 112 1
 552 + 111 11111 1111 11 1 1 1
```

```
2546 +3 333 3 3 3 3   3 333 33
 606 +3333 3 333333 3333
 150 +3333 3 333333333333  3 3
1256 +3333 3 233333333333  1
1140 +3333 2 333333333333    1
1408 +3332 2 233332233322 2  3
 327 +3 333 3 3 2 3   3 32  21
1352 +3 331 3 3 3 3   1 31  11
 695 +3 1 2 3 3 3 3   3 311 12
 893 +2232 3 23 23323222     1
 474 +2222 2 223332333222 12 2
 943 +3 232 2 2 3 3   1 31  11
1327 +3312 2 22 23  22322    1
  42 +223  2 122232232112  1 1
1558 +2323 1 323323212211  1 1
1640 +3231 3 112111211112  1 2
 404 +2222   2212  2212 1 11 1
2264 +2 112 2 1 1 1   2 112 11
 284 +1 212 1 2 2 1   1 11  11
2463 +2 211 1 1 1 1   2 111 11
 687 +1 211 1 1 1 1   1 11  11
```

```
 134 + 33333 3 3 33    3333
 509 + 33333 3 3 33    3333
1915 + 33333 3 3 33    2 33
1090 + 3333  3 3 33    3 32
 793 +3 33  3 2 2  3333   2233
2387 + 22223 3 3 23    3222
2338 + 33333 3 3 33    2321
2255 +      3   3     2   2
 905 +3 23  3 2 3   233   1222
2228 +3 33  2 3     3 3   2111
 902 +2 33  3 2 2  2222   3132
1597 + 2 2 2 3 1 12       12
 211 + 22222 2 2 22      22
 198 +3 33  2 1 2   311   1 11
 153 +2  2  2 2 3   121   11 1
 875 +2 22  2 2 2  2 12   2211
  65 + 22222 2 2 11    2121
2098 +2 21  2 1 1  2211   1112
2524 +2 21  1 2 2    12   11
1873 + 12111 1 2 11    1111
1772 + 11111 1 1 1     1111
```

Fig. 2. Three samples of different respondents answering different items

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

The data shown in Fig. 2 were extracted from the same larger data set of over 2,500 cases and 78 items as the data shown in Fig. 1. The total data set's measurement reliabilities ranged from 0.94 to 0.97. Copying the entire file three times, and then removing different rows and columns from each of the three different subsamples, as shown in Fig. 2, conveys the point in concrete terms that there is a real construct here that asserts itself as an independent entity that separates from the local contingencies of its origin. Different sets of items can measure the same thing across different samples of people in the same unit.

3. Formally stating the model

The relevant Rasch model for this kind of rating scale data [34-35] is:

$$\log\left(\frac{\pi_{nix}}{\pi_{ni(x-1)}}\right) = \beta_n - (\delta_i + \tau_x) \tag{1}$$

where the log is base e and the rest of the left side of the equation is the probability π of person n being observed responding to item i in category x-1 relative to category x. This specification of the log-odds structure required in the data is evaluated relative to the difference between the ability β of person n and the combined difficulties δ and τ of

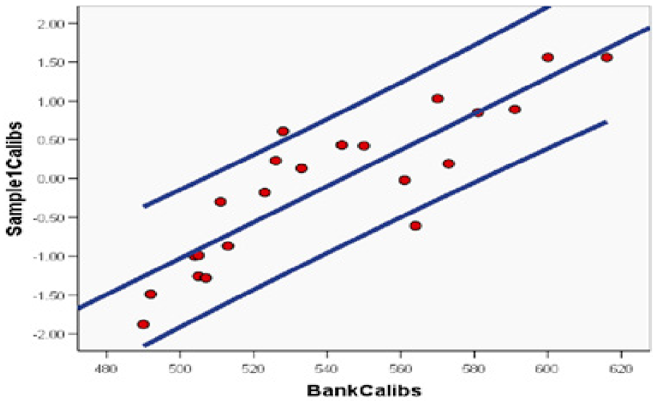

Fig. 3. Sample 1 (a) calibrations vs overall bank calibrations from total sample of over 2,500 cases

item i and category x. When the observed scores function as sufficient statistics [3], the data approximate this model, and the parameters are separable, in the sense that item calibrations are independent of the person measures, and vice versa.

The three data sets in Fig. 1 were each fit to the same Rasch model using the Winsteps software [32]. Two of the 24 items had only 3-12 responses each, far fewer than the 30 that have been shown typically sufficient for reproducing difficulty order schemas [36]. The other 22 items have an average of 40 responses each across the three data sets. The logit difficulty estimates for these items correlate .82, .88, and .90; the correlations are 1.00 after disattenuation [37]. The three subsample sets of calibrations correlate .88, .90, and .91 with the calibrations from the total sample of over 2,500 cases; again, the correlations are 1.00 after disattenuation.

Furthermore, the calibrations do not simply stay in the same *order* across samples, they retain the same unit size; i.e., they are invariantly positioned relative to one another, within the range of error. Fig. 3 shows the scatter plot of the Fig. 1 (a) calibrations vs the overall item bank calibration anchors. The confidence intervals are fairly wide, as is expected given the error and reliabilities obtained with Sample 1's per-item sample sizes of about 40. The confidence intervals become progressively narrower as sample size and reliability increase [38].

Taking up Taylor's [1, pp. 166-168] language again, the regularity in the flux of data identified in Fig.1 and Fig. 2 suggests the form of a schema (Fig. 4) that will enable the system to recognize the pattern if it is encountered again. The second moment in complex adaptive functionality is obtained when the data are compressed via an algorithm that enables the system to recognize the pattern again, and unfold its details efficiently. The more reducible the data are, the greater the efficiency obtained in the system's functioning. Compression is arguably most efficient when a data reduction algorithm is able to make use of sufficient statistics. Rasch models, for which raw scores are minimally sufficient statistics (both necessary and sufficient) [3], enable information processing that not only filters noise, but capitalizes on the information in it, and so appears to create order from chaos.

Thus the chaos may be more mere appearance than actual pure noise. At this point in his description of the five features of complex adaptive systems, Taylor [1, p. 166] notes that “the schemata or algorithms in complex adaptive systems are emergent and can change,“ and so, he suggests, they cannot be preprogrammed or fixed. But the appearance of change may be one thing when viewed in the absolute terms of concrete facts and events (Figs. 1-2), and another thing when viewed stochastically in terms of log-odds units bound by uncertainty. Adaptive, emergent, and changeable algorithms are routinely preprogrammed and fixed, as is suggested by the nature of the patterns shown in Fig. 1-3. This speaks to the more important, superceding point raised by Taylor [1, p. 167] that complex adaptive systems, operating as they do by processing information, require the filtering of noise, and the capacity to create order from it. Adaptation may well and often require changes to the concrete terms of the actual information processed, but that concrete observation model may nonetheless fit well within the parameters of the overarching abstract idealized model. The same compression scheme may then still serve as an effective strategy in conserving memory resources.

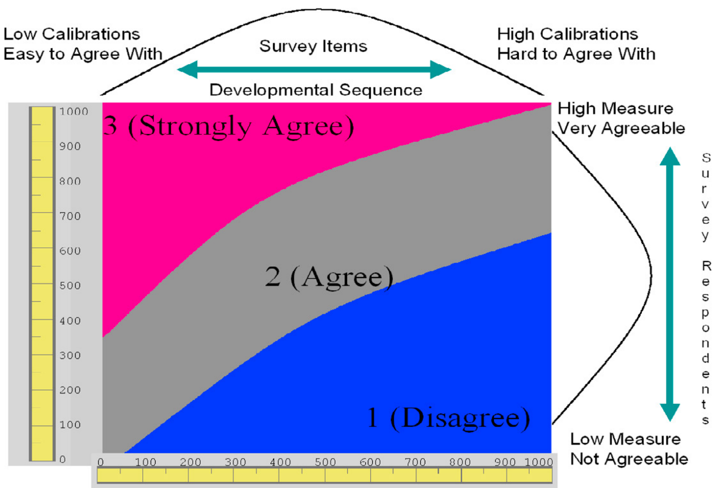

Fig. 4. The modeled pattern expected from the data in Fig. 1 and Fig. 2

Given the wide range of Rasch model applications over the last 30 years and more, patterns like this seem to be pervasive across the domains of education, health care, human resources, social services, natural resource management, etc. Given that test, survey, and assessment data organize themselves in these kinds of repeatable and reproducible patterns, significant opportunities present themselves for constructing and capitalizing on the efficiencies to be gained from capitalizing on the predictability of the compression schemes. But few are looking for these patterns, and those who are do not seem to realize their importance, or don't know how to tell their story.

4. Discussion

The coherence with which things speak to us across data sets does not, of course, happen by itself. Questions have to be framed properly. Their composition and phrasing has to be informed by a theory of what counts as less and more of the thing to be measured [5-6]. The preparation of a medium in which a self-organizing conceptual entity like a measured construct can manifest itself is akin to laboratory work in which a micro-environment for a form of life is created. What we do, in effect, is pose questions to people likely to be caught up in the play of the intended construct's language game. In the back and forth of the question and answer process, everyone involved participates so far as they are able in the construct's self-representative activity. The thing itself then is able to write its story on the abstract conceptual medium provided to it, acting through the participants in the dialogue.

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

> On the basis of [one of the equalities implied in Rasch's model] we may estimate the item parameters independently of the personal parameters, the latter having been replaced by something observable, namely, by the individual total number of correct answers. Furthermore, on the basis of [the next equation] we may estimate the personal parameters without knowing the item parameters which have been replaced by the total number of correct answers per item. Finally, [the third equation] allows for checks on the model [another equation] which are independent of all the parameters, relying only on the observations" [39, p. 325; 2, p. 122].

Fig. 1 illustrates Rasch's first sentence here, in that they show that the same data pattern forming across samples for a group of items. Because the pattern stays the same across the observed responses, we can estimate the calibration parameters with no concern for the particular persons measured. Then, because the pattern in the data stays the same across groups of items for a given sample of people, we can estimate measures for them without knowing what the item parameters are. Finally, as is illustrated in Fig. 4, we can evaluate the clarity of our thinking, and the model, by examining the observations alone, apart from any concern with the particulars of who answered the questions or which questions were asked.

5. Practical implications

The stability of item calibrations across samples, and of measures across item samples, has supported the emergence of a testing industry that capitalizes on the repeatable confirmation of predictable data patterns. Study of these kinds of patterns leads to opportunities for

- Unit standards [10-14]
  - All instruments measuring same construct are equated to common metrics
  - Measures interpretable within students over time, and across students, classrooms, schools, etc.
- Data never fit the model but this may not detract from pragmatic utility [2]
  - Major goal is to reveal actionable anomalies [31, 40]
  - Kidmaps inform clinical and instructional planning [41-44]
- Mass customization [15]
  - Items individually tailored to respondents' needs
  - All measures expressed in single common metric
- Data quality assessment [40]

  - Internal consistency evaluation
  - Goes beyond checking codes
- Data volume reduction [3]
  - One interval measure per person, unit, division, facility
  - Instead of multiple ordinal numbers for each per item
- Meaningful measures [5,6,45]
  - A substantial thing measured that adds up the way the numbers do
  - Not numbers that vary in meaning relative to irrelevant factors
- Fairer measures [46]
  - Data from judges rating performances are evaluated for internal consistency
  - Individual measures are adjusted across raters, creating a level playing field
- Instructionally-embedded assessments *for* learning
  - Formative assessments that maximize learning outcomes [47]
  - Informational coherence across multilevel discontinuities [9, 17-21]

Efforts to date to apply agent-based models in evaluation or education research [48] tend to overlook basic measurement principles. Conversely, research on environmental education and on educational environments informed by advanced measurement models [49-50] neglects opportunities for modeling and measuring ecosystem dynamics in actionable ways capable of informing policy and local decision processes. It then happens that individual-level measurement models of environmental attitudes and behaviors [51] map substantively and scientifically meaningful units of measurement, but focus on cross-sectional, local, single-level and static slices in time. At the same time, individual-level modeling in the context of complex adaptive systems and ecologies [7-8,48], in contrast, focuses on dynamically evolving interactions but fails to capitalize on relevant distinctions between statistical and scientific approaches to data, theory, and instruments [52]. There is a great need for this work to leverage recent and longstanding advances in the scientific quality of individual-level measurement in psychology and the social sciences [2-6,10-21,25-26,31-36,38-46,52-58] offering inferentially separable model parameters, minimally sufficient statistics in estimation, theory-informed instrument design, and experimental tests of the hypothesis of an additive unit.

Agent-based models [7-8] should be integrated with advanced multi-unidimensional, multifaceted, multilevel, and growth variations on stochastic measurement models and methods [44,53-58]. These features of precision measurement could be implemented as decision supports in broad social ecologies composed of networks of various stakeholders pursuing separate but related interests in different facets of ostensibly the same boundary objects. Partially harmonized and partially dissonant individual behavior analogues could then be coordinated across stakeholder groups [31] relative to metrological standards [10-14] in generative research and practice more likely to give rise to viral wisdom of crowds phenomena than currently available methods allow. This kind of modeling research will be useful in promoting sustainable and self-organizing ecosystems of human, social, and natural capital in economies redefining profit in terms of authentic wealth and value for life [59-60]. The interrelations of science and the economy [61] are such that the capacity to scientifically measure meaningful amounts in a distributed cognitive system enables the commercial exchange of value. As social and economic flows of complex adaptive functionality in education, health care, social services, etc. are better understood, a new science of self-organizing living systems will bring psychology and the social sciences together with the natural sciences [62-64]. Though the challenges are huge, the consequences for shifting the culture away from crass commercialism toward the production of authentic wealth and the end of inflationary spirals in education, government, and health care may be profound [10,31].

References

[1] Taylor, Mark. (2003) *The moment of complexity: Emerging network culture*. Chicago, University of Chicago Press

[2] Rasch, Georg. (1960) *Probabilistic models for some intelligence and attainment tests* (Reprint, Chicago, University of Chicago Press, 1980). Copenhagen, Danmarks Paedogogiske Institut

[3] Andrich, David. (2010) "Sufficiency and conditional estimation of person parameters in the polytomous Rasch model." *Psychometrika* **75** (**2**): 292-308.

[4] Bond, Trevor, and Christine Fox. (2015) *Applying the Rasch model: Fundamental measurement in the human sciences, 3d edition*. New York, Routledge

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

[6] Stenner, A. Jackson, William P. Fisher, Jr., Mark H. Stone, and Donald S. Burdick. (2013) "Causal Rasch models." *Frontiers in Psychology: Quantitative Psychology and Measurement* **4** (**536**): 1-14.

[7] Railsback, Steven F. (2001) "Concepts from complex adaptive systems as a framework for individual-based modelling." *Ecological Modelling* **139** (**1**): 47-62.

[8] Grimm, Volker, and Steven F. Railsback. (2013) *Individual-based modeling and ecology*. Princeton, NJ, Princeton University Press

[9] Star, Susan L., and Karen Ruhleder. (1996) "Steps toward an ecology of infrastructure: Design and access for large information spaces." *Information Systems Research*, **7** (**1**): 111-134.

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

[11] Mari, Luca, and Mark Wilson. (2014) "An introduction to the Rasch measurement approach for metrologists." *Measurement* **51**: 315-327.

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

[13] Pendrill, Leslie, and William P. Fisher, Jr. (2015) "Counting and quantification: Comparing psychometric and metrological perspectives on visual perceptions of number." *Measurement* **71**: 46-55.

[14] Wilson, Mark R. (2013) "Using the concept of a measurement system to characterize measurement models used in psychometrics." *Measurement* **46**: 3766-3774.

[15] Barney, Matthew, & William P. Fisher, Jr. (2016) "Adaptive measurement and assessment." *Annual Review of Organizational Psychology and Organizational Behavior* **3**: 469-490.

[16] Fisher, William P., Jr., and A. Jackson Stenner. (2016) "Theory-based metrological traceability in education: A reading measurement network." *Measurement* **92:** 489-496.

[17] Fisher, William P., Jr., and A. Jackson Stenner. (2017) "Ecologizing vs modernizing in measurement and metrology." *Journal of Physics Conference Series* [in press].

[18] Fisher, William P., Jr., Emily Pey-Tee Oon,, and Spencer Benson. (2017) "Applying Design Thinking to systemic problems in educational assessment information management." *Journal of Physics Conference Series* [in press].

[19] Gorin, Joanna S., and Robert Mislevy. (2013) *Inherent measurement challenges in the next generation science standards for both formative and summative assessment* (K-12 Center at Educational Testing Service, Invitational Research Symposium on Science Assessment). Princeton, NJ, ETS. Retrieved from http://www.ets.org/Media/Research/pdf/gorin-mislevy.pdf

[20] National Research Council. (2006) *Systems for State Science Assessment*. Committee on Test Design for K-12 Science Achievement. M. R. Wilson and M. W. Bertenthal, (Eds.). Board on Testing and Assessment, Center for Education. Division of Behavioral and Social Sciences and Education. Washington, DC, The National Academies Press

[21] Wilson, Mark. (Ed.). (2004) *National Society for the Study of Education Yearbooks. Vol. 103, Part II: Towards coherence between classroom assessment and accountability*. Chicago, University of Chicago Press

[22] Matsumoto, K., and I. Tsuda. (1983) "Noise-induced order." *Journal of Statistical Physics* **31** (**1**), 87-106.

[23] Repperger, D. W., and K. A. Farris. (2010) "Stochastic resonance —a nonlinear control theory interpretation." *International Journal of Systems Science* **41** (**7**), 897-907.

[24] Galison, Peter. (1997) *Image and logic: A material culture of microphysics*. Chicago, University of Chicago Press

[25] Fisher, William P., Jr., and Mark Wilson. (2015) "Building a productive trading zone in educational assessment research and practice." *Pensamiento Educativo: Revista de Investigacion Educacional Latinoamericana* **52** (**2**), 55-78.

[26] Andrich, David. (2017) "A law of ordinal random error: The Rasch measurement model and random error distributions of ordinal assessments." *Journal of Physics Conference Series* in press.

[27] Nersessian, Nancy J. (2008) *Creating scientific concepts*. Cambridge, Massachusetts, MIT Press

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

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

[30] Weichhart, George, Thomas Feiner, and Chris Stary. (2010) "Implementing organisational interoperability--The SUddEN approach." *Computers in Industry* **61**: 152-160.

[31] Fisher, William P., Jr. (2011) "Bringing human, social, and natural capital to life: Practical consequences and opportunities." *Journal of Applied Measurement* **12** (**1**), 49-66.

[32] Linacre, John M. (2016) *WINSTEPS Rasch-Model computer program, v. 3.92.0*. Chicago, Winsteps.com

[33] Elbaum, Batya, William P. Fisher, Jr., and W. Alan Coulter. (2011) "Measuring schools' efforts to partner with parents of children served under IDEA: Scaling and standard setting for accountability reporting." *Journal of Applied Measurement* **12** (**3**): 261-278.

[34] Andrich, David. (1978) "A rating formulation for ordered response categories." *Psychometrika* **43** (**4**): 561-573.

[35] Wright, Benjamin D., and Geofferey N. Masters. (1982) *Rating scale analysis: Rasch measurement*. Chicago, MESA Press

[36] Green, Kathy E., and Richard M. Smith. (1987) "A comparison of two methods of decomposing item difficulties." *Journal of Educational Statistics* **12** (**4**): 369-381.

[37] Muchinsky, P. M. (1996) "The correction for attenuation." *Educational and Psychological Measurement* **56** (**1**): 63-75.

[38] Fisher, William. P., Jr., Batya Elbaum, and W. Alan Coulter. (2010) "Reliability, precision, and measurement in the context of data from ability tests, surveys, and assessments." *Journal of Physics, Conference Series* **238** (**1**), 012036.

[39] Rasch, Georg. (1961) "On general laws and the meaning of measurement in psychology." In J. Neyman (Ed.), *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability: Volume IV: Contributions to biology and problems of medicine* (pp. 321-333). Berkeley, California, University of California Press

[40] Smith, Richard M. (2000) "Fit analysis in latent trait measurement models." *Journal of Applied Measurement* **1** (**2**): 199-218.

[41] Chien, Tsair-Wei, John M. Linacre, and Wen-Chung Wang. (2011) "Examining student ability using KIDMAP fit statistics of Rasch analysis in Excel." In *Communications in Computer and Information Science: Vol. 201. Advances in Information Technology and Education* (pp. 578-585). Berlin, Springer Verlag

[42] Chien, Tsair-Wei, Yang Shao, and Shu-Chun Kuo. (2017) "Development of a Microsoft Excel tool for one-parameter Rasch model of continuous items: An application to a safety attitude survey." *BMC Medical Research Methodology* **17** (**4**).

[43] Mead, Ronald J. (2009) "The ISR: Intelligent Student Reports." *Journal of Applied Measurement* **10** (**2**), 208-224.

[44] Wright, Benjamin D., Ronald J. Mead, and Larry H. Ludlow. (1980) *KIDMAP: person-by-item interaction mapping* (Tech. Rep., MESA Memorandum #29). Chicago, MESA Press [http://www.rasch.org/memo29.pdf].

[45] Stone, M. H., Benjamin D. Wright, and A. Jackson Stenner. (1999) "Mapping variables." *Journal of Outcome Measurement* **3** (**4**): 308-322.

[46] Linacre, John M., George Engelhard, Donna Surges Tatum, and Carol M. Myford. (1994) "Measurement with judges: Many-faceted conjoint measurement." *International Journal of Educational Research* **21** (**6**): 569-577.

[47] Hattie, John. (2008) *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* New York, Routledge

[48] Morell, J. A., R. Hilscher, S. Magura, and J. Ford. (2010) "Integrating evaluation and agent-based modeling: rationale and an example for adopting evidence-based practices." *Journal of MultiDisciplinary Evaluation* **6** (**14**): 32-57.

[49] Cavanagh, Robert F. (2015) "A unified model of student engagement in classroom learning and classroom learning environment: one measure and one underlying construct." *Learning Environments Research* **18** (**3**): 349-361.

[50] Frick, J., Florian G. Kaiser, and Mark Wilson. (2004) "Environmental knowledge and conservation behavior: Exploring prevalence and structure in a representative sample." *Personality and Individual Differences* **37**: 1597-1613.

[51] Roczen, N., Florian G. Kaiser, F. X. Bogner, and Mark Wilson. (2014) "A competence model for environmental education." *Environment and Behavior* **46** (**8**): 972-992.

[52] Wilson, Mark. (2013) "Seeking a balance between the statistical and scientific elements in psychometrics." *Psychometrika* **78 (2)**: 211-236.

[53] Adams, Raymond J., Mark Wilson, and Wen-Chung Wang. (1997) "The Multidimensional Random Coefficients Multinomial Logit Model." *Applied Psychological Measurement* **21** (**1**): 1-23.

[54] Adams, Raymond J., Mark Wilson, and Margaret Wu. (1997) "Multilevel item response theory models: An approach to errors in variables of regression." *Journal of Educational and Behavioral Statistics* **22:** 47-76.

[55] Beretvas, N., and Akhito Kamata. (2007) "Part II. Multi-level measurement Rasch models." In E. V. Smith, Jr. & R. M. Smith (Eds.), *Rasch measurement: Advanced and specialized applications*. Maple Grove, Minnesota, JAM Press.

[56] Briggs, Derek, and Mark Wilson. (2003) "An introduction to multidimensional measurement using Rasch models." *Journal of Applied Measurement* **4**: 87-100.

[57] Fischer, Gerhard H. (1973) "The linear logistic test model as an instrument in educational research." *Acta Psychologica* **37**: 359-374.

[58] Williamson, Gary L., Jill Fitzgerald, and A. Jackson Stenner. (2014) "Student reading growth illuminates the common core text-complexity standard." *The Elementary School Journal* **115** (**2**), 230-254.

[59] Ekins, Paul. (1992) "A four-capital model of wealth creation." In P. Ekins & M. Max-Neef (Eds.), *Real-life economics: Understanding wealth creation* (pp. 147-155). London, Routledge

[60] Ekins, Paul, S. Dresner, and K. Dahlstrom. (2008) "The four-capital method of sustainable development evaluation." *European Environment* **18** (**2**): 63-80.

[61] Ashworth, William J. (2004) "Metrology and the state: Science, revenue, and commerce." *Science* **306** (**5700**), 1314-1317.

[62] Chakraborti, A., I. M. Toke, M. Patriarca, and F. Abergel. (2011) "Econophysics review: I. Empirical facts." *Quantitative Finance* **11** (**7**): 991-1012.

[63] Colander, D. (2014) "Capitalism as a complex evolving system." *Ekonomi-tek-International Economics Journal* **3** (**1**), 13-22.

[64] Colander, D., and R. Kupers. (2014) *Complexity and the art of public policy: Solving society's problems from the bottom up*. Princeton, New Jersey, Princeton University Press.