

PAPER • OPEN ACCESS

Ecologizing vs modernizing in measurement and metrology

To cite this article: W P Fisher Jr and A J Stenner 2018 *J. Phys.: Conf. Ser.* **1044** 012025

View the [article online](#) for updates and enhancements.

Related content

- [Defining and Measuring Nature: Measurement in the Early Modern Period](#)
J H Williams
- [Metrology in Pharmaceutical Industry - A Case Study](#)
Priscila D. Yuvamoto, Ricardo K. S. Ferraam and Elizabeth S. Nascimento
- [Forensic Metrology: Its Importance and Evolution in the United States](#)
JD Ted Vosk

Ecologizing vs modernizing in measurement and metrology

W P Fisher Jr¹ and A J Stenner²

¹ Research Associate, BEAR Center, Graduate School of Education, University of California, Berkeley, Berkeley California 94702 USA

² Senior Scientist, MetaMetrics, Inc., Durham North Carolina 27713 USA

¹E-mail: wfisher@berkeley.edu

²E-mail: jstenner@lexile.com

Abstract. The complex multilevel nature of language is a significant barrier to the conceptualization, design, and implementation of effective large scale measurement communications systems in education. The barrier's existence and challenges are obscured, first, by the modern emphasis on data as the hallmark criterion of an objective basis for decision making. The situation is further obscured by the postmodern observation that attention is focused by linguistic, historical, and cultural concerns. Though it is true that data are salient only to the extent implicit or explicit theory is brought to bear, postmodernism does not follow through with its insights into how to create effective information infrastructures. Data, theory, and instruments function interdependently in the historical successes of science to bring things into language. The ecological economy of language is realized by the efficiencies obtained when new concepts produced by collective learning processes are embodied in tools distributed throughout cognitive ecosystems. These ecosystems are comprised of discontinuous but manageably coherent levels of increasingly complex forms of information, following the pattern of everyday language. An emerging reading measurement ecosystem of this kind offers a developmentally, horizontally, and vertically coherent language for managing literacy education across niches defined by classrooms, schools, districts, homes, offices, libraries, test and assessment agencies, and book publishers.

1. Introduction

The fact that language functions at varying levels of complexity and abstraction has long been recognized and addressed by philosophers and linguists [1-4]. Only recently, however, have information systems designers begun to address the conceptual and practical problems associated with the multilevel nature of language [5-11]. The problems encountered are particularly acute in the context of education's need for developmental, horizontal, and vertical coherence across the multiple kinds of assessments employed [12-14]. Many problems encountered in the design of educational measurement information systems are rooted in the unrecognized and unaddressed complexities of language's multiple levels of abstraction. Until these problems are confronted, little in the way of effective measurement communications in education and other fields can be expected.

Questions that arise in this context ask if it possible to reconcile the tension between the need in education and other areas, such as health care, for local, personal, and customized assessment

¹ To whom any correspondence should be addressed.



technologies, on the one hand, and the need for those technologies to be traceable to common standards. Must standards always require an imposition of irrelevant or misguided criteria on local idiosyncracies? Will the culture wars between modern universalism and postmodern relativism become a permanent feature of intellectual life? Or will embedded relationality prove its capacity to integrate the opposition of modern universalism and postmodern relativism? Is there some way in which we can care for our educational and healthcare measurement technologies as we do our children [15], with the aim of preventing the creation of new well-intentioned but nonetheless monstrous effects? Can the unique particulars of local experience be meaningfully and usefully integrated into our institutions' information infrastructures [6]? Is it possible for the values and experiences of individuals to be incorporated deeply and authentically into planned, built, and legislated forms of social life? Could the skills, knowledge, and responsibilities of those individuals be enhanced by their participation in those forms of social life?

Positive answers to these questions are suggested by recent developments in philosophy and the history and social studies of science [9-11, 15-19], and in measurement theory synthesizing engineering and psychometric perspectives [20-24].

Because the multilevel complexities of language are not taken into account, education, health care, human resource management, social services, and many other areas of life are marked by a kind of schizophrenia [4-5] that emerges in terms of the dissonance between a caring focus on individual needs for learning and healing, on the one hand, and demands for accountability focused on standards and comparability. Support for irrevocable concerns with the individual student's and patient's spontaneous processes of development and healing stands as an immovable thesis that is increasingly in opposition to the larger social antithesis of an imposed demand for evidence proving the achievement of quality standards. How might new institutional forms of social life resolving the schizophrenic break be formed at a higher level of information infrastructure complexity? How might those measured forms of life synthetically integrate the necessary concern for development and healing with a new ecologizing bottom-up approach to accountability and standards that authentically embodies individual uniqueness?

Linguistic communication systems incorporate within-individual processes distinct from, but interacting with, mid-level processes between individuals, and which in turn are distinct from but interacting with high-level group processes. These levels of complexity in communication have informed practical applications in epidemiology [7] leading to new, productive relationships between clinical medicine and public health efforts [8].

Might not a similar kind of productivity be possible in education if we apply similar ecological approaches to conceiving, designing, and implementing developmental, horizontal, and vertical coherence [12-14] in educational assessment? Cognitive and social ecologies contextualize problems of coherence in educational assessment in ways that may be key to understanding learning in each distinct niche of the various environments in which it lives. Learning varies across these levels of complexity in ways that cannot be grasped directly from individual measures. What form might conceivably be taken by educational assessment communications systems capable of supporting broad-scale efforts at sorting out the sources of distinct classes of effects on learning?

2. Steps toward an ecology of assessment information infrastructure

What specific problems are encountered when discontinuities in levels of linguistic abstraction are ignored in the design of information infrastructures? Each different paradigmatic approach to understanding the relation of data, theory, and instruments—modern positivist, postmodern antipositivist, and unmodern postpositivist [25]—creates its own kinds of problems in this regard. Philosophically modernist conceptions of science are positivist in the sense of prioritizing a focus on data as the ultimate criterion of objectivity. Postmodernism, in contrast, is sensitized by historical changes in what data count as worthy of attention and so is concerned with the role of theory in making data salient. Unmodern (also known as amodern) postpositivist perspectives [17, 26] assert that the debate between modern and postmodern focuses too exclusively on the mutual implication of theory and data, and so will remain unresolved as long as the roles of instruments and knowledge technologies are not taken into account. Instruments encapsulate what is learned from data and what

can be explained by theory. That assemblage is then circulated within networks sharing common standards (vocabularies, units, grammars, etc.).

Unmodern philosophical perspectives and research in the history of science [5, 9-11, 15-19] focus on the collective cognition and team-based coordinations made possible when theory-data-instrument assemblages are expressed in a uniform language distributed throughout a community of practice. Metrology's concern with measuring instruments traceable to unit standards then becomes a matter of focal interest as a way in which everyday model-based reasoning has been extended productively into science [9]. Recent developments suggesting metrological paths forward for the constructs of psychology and the social sciences [20-24, 27-28] also extend everyday model-based reasoning [27] and open up new possibilities for enhanced innovation in education, health care and other fields. A significant problem that remains unaddressed is how varying levels of information complexity can be integrated into a new metrological culture encompassing all of the arts and sciences.

For instance, in applying educational measures for formative and accountability purposes, we face a problem of coherently coordinating and aligning representations of processes and outcomes across discontinuous levels of complexity. Taking steps toward an ecology of infrastructure alongside Bateson's [4] steps to an ecology of mind, Star and Ruhleder [5] point out:

These discontinuities have the same conceptual importance for the relationship between information infrastructure and organizational transformation that Bateson's work on the double bind had for the psychology of schizophrenia. If we, in large-scale information systems implementation, design messaging systems blind to the discontinuous nature of the different levels of context, we end up with organizations which are split and confused, systems which are unused or circumvented, and a set of circumstances of our own creation which more deeply impress disparities on the organizational landscape.

In educational assessment, the usual approach to counting up correct answers and treating that score as a measure is known as True Score Theory (TST) [28]. TST works at the local level of simple denotation. Bateson's [4] example is "The cat is on the mat." TST says, "Your test score is x ." Bateson observes that this kind of simple statement of fact is very different from the metalinguistic statement, "The word 'cat' has no fur." The object of reference has changed from a small mammal to a word. Similarly, in saying "Your score of x on the test means you're proficient," TST wants to shift the focus from recorded responses to a broader linguistic construction. But it does so in a limited way.

The difference is that a word like 'cat' has a place in the larger culture recognized by virtually everyone who can speak the language, whereas the meaning of the test score depends on the local context of particular questions asked. 'Cat' emerged from a self-organized historical process that was not directed by any one person or group, and that resulted in a commonly shared sign universally understood within a cultural context as representing a particular kind of animal. The TST score, in contrast, was determined by the intentions of someone who composed the questions and recorded the responses, not by a collective, undirected, self-organized process. Statistical models applied to these data take a modern positivist approach primarily interested in describing the facts of the situation, with little or no effort invested in developing construct theories or in calibrating interval-level measuring instruments capable of supporting more complex metalinguistic statements.

What possibilities exist in educational measurement for fostering the manifestation and observation of emergent, self-organizing phenomena that could be represented in signs referring to shared objects in universally accessible and shared symbols? Do not group level constructs emerge from the internal coherence of responses to test questions? Psychometric constructs are invariant in a way that is not under the control of intentions imposed from the outside by the person asking the questions and recording the responses (though the construct is given a medium for its self-organized expression). Might not there be a way to interpret and contextualize these constructs so as to bring them into language in meaningful and useful terms, in a way more akin to the way the word "cat" emerged than is obtained by TST?

Rasch's models for unidimensional measurement [28-31] are sensitive to the distinction between the local denotative and mid-range metalinguistic levels. These models enable the development of construct theories and the calibration of interval-level instruments. Stopping with analyses applying Rasch's models for measurement to given data sets, though, does not complete the process. To follow through, we need to distribute the concept throughout a network of shared signification by embodying it in words that consistently mean the same thing wherever they turn up. This is the task of metrological traceability. Coordinating and aligning meanings on this scale is incredibly difficult given the variation in the questions and answers that exists, and, especially, given the widespread lack of appreciation for the value of explanatory, predictive theory capable of efficiently removing the need for data-based measure estimation [32-33].

The difficulty becomes even greater at the third, metacommunicative level. This level's statements include "My telling you where to find the cat was friendly," or "Your instructional methods in the class with the mean test score that changed so much must have been superior." Now comments on associated aspects of the situation are being brought to bear and have to be evaluated. Maybe saying where to find the cat was not friendly but was motivated by a desire to have the animal removed from your garden. Perhaps the mean test score changed less because of instructional methods than because of the students' home environments, genetic propensities for learning, or some combination of these.

So, in the same way that a group-level effect occurs in the shift from denotative to metalinguistic communications, another one occurs at the metacommunicative level. Here we encounter the problem of the ecological fallacy, the problem of inferring individual-level characteristics from group-level statistics. We overcome one facet of this problem with Rasch's probabilistic models by testing the hypothesis of meaningful comparability at the group-level of coherent constructs, and by accepting individual departures from the model expectations in technologies like the kidmap [34]. But other factors affecting comparability remain unevaluated.

That is, the causal relationship between question and answer exists in a context in which the consistency of ability relative to difficulty may be stable across a wide range of different circumstances, such as nutrition, genetics, environmental quality, etc. Until we identify what these factors are and account for them in a metrological context in which we share common languages and can sort things out effectively in generalizable ways, we will continue to fail in our efforts to create institutions that systematically support and value human sensitivity and relationships. Models and methods effectively integrating the denotative, metalinguistic and metacommunicative levels [35] take these hierarchical effects into account in ways that do not penalize but reward enacting these qualities in individual and small group relationships in classrooms and clinics.

3. A reading measurement ecosystem

Reading measures are linked together in an ecosystem that has capitalized on the literacy form of life that consistently asserts itself across samples of students, texts, test items, time, and space [24]. Niches in this ecosystem span a wide range of classrooms, schools, homes, libraries, testing agencies, and book publishers. Reading test item difficulties have been shown to be remarkably stable over decades of use [36] and moreover can be predicted by an explanatory theory accounting for over 90 percent of the observed variance [24, 32]. More than 100 English language reading tests across the world measure in a common unit. Over 30 million student measures annually are interpreted relative to 250,000 book measures and 200 million article measures, where matching student and text measures predict a 75 percent comprehension rate. Books, articles, assessments, and students have been brought into a common frame of reference in a process now over 27 years old and still accelerating. Text complexity corresponds with reading learning progressions such that student measures enable the individualization of instruction.

Each of the three forms of coherence sought for educational assessment information infrastructures [13-15] is provided in this system. Student measures are tracked over time and across grade levels, instantiating developmental coherence. Teachers are able to compare learning outcomes across their own and each other's classes, realizing horizontal coherence. And in many locations, state end-of-year

or graduation tests report in the common unit, providing parents, students, teachers, principals, librarians, researchers, and the public with the vertical coherence needed for connecting classroom formative assessments with accountability standards.

4. Implications

An obvious starting point for developing effective educational assessment information infrastructures and associated ecosystem communities of practice lies in the complexity of the partially connected relationships between stakeholder groups in technologically mediated networks of various kinds. Galison [25, pp. 844-845] wonders what new analogy might be capable of informing models of science that have emerged in the wake of postmodern deconstructions showing communities of theoreticians, experimentalists, and instrumentalists as disunified. Galison raises examples of amorphous crystals in electronics and laminated materials in structural engineering as instances in which certain amounts of disorder provide greater reliability and utility than could be obtained if the materials were perfectly and rigidly structured. Wittgenstein's metaphor of concepts as intertwined strands of fibre in a thread that is stronger than it would be if it were formed from only one single continuous fibre is plainly relevant, as Galison recognizes. But Galison is sensitive to the dynamic processes involved in social networks and seeks instead a nonmechanical way of coordinating the "different symbolic and material actions [through which] people create the binding culture of science." Berg and Timmermans [10, p. 56] independently concur, saying that the stability and reach of the medical decision networks they studied were "not due to more (precise) instructions: the protocol's logistics could thrive only by parasitically drawing upon its own disorder." The ways in which differing orders of things chaotically aggregate in harmonic and dissonant patterns is similar to the effects of stochastic resonance, which may provide an apt metaphor for psychological and social ecosystem processes [37].

Instead of demanding strict conformity with standards, then, it is likely more realistic and productive to think of standards as potential universals in partially interconnected, resonant, and multilevel traceability networks, bypassing the strictly local problem-solution dependency and the universal problem-solution independence at the same time [11, p. 229]. These "glocal" media, simultaneously local and global, are characterized by Ricoeur [38, p. 289] as potential or inchoate universals. Dewey [39, p. 215] similarly held that "The local is the ultimate universal, and as near an absolute as exists." The interconnections of metrological networks supporting local approximations and translations of standards is pointed to by Golinski [18, p. 35] as replacing the uniform universality assumed in modern science. And Haraway [16, pp. 439-440] suggests another account as to how locally embedded relationships offer an alternative to both relativism and transcendence.

The sustainability opportunities created within an ecologizing paradigm stem from the co-evolution of (a) concepts embodied in linguistic and measurement technologies and (b) the institutional rules, roles, and responsibilities within multilevel social, political, and economic ecologies [40]. The end results are systems of tools embodying individually unique problem-solution unities that are useful in negotiating local particularities while still recognizable as belonging to an identifiable general class. These results suggest potentially large payoffs of new analogies from existing online engineering models of global cooperation enabling intelligent metrology applications [41]. Perhaps caring for our measuring technologies in education and other fields the same way we care for our children will yet lead to creation of forms of social life sensitive to the values and experiences of those who inhabit them.

References

- [1] Whitehead A N and Russell B 1910-1913 *Principia Mathematica*, 3 vols (Cambridge, England: Cambridge University Press)
- [2] Wittgenstein L 1922 *Tractatus logico-philosophicus* (London: Harcourt Brace)
- [3] Carnap R 2003 *The logical syntax of language* (New York: Harcourt Brace)
- [4] Bateson G 1972 *Steps to an ecology of mind* (Chicago: University of Chicago Press)
- [5] Star S L and Ruhleder K 1996 *Information Systems Research* **7** 111-34
- [6] Scott J C 1998 *Seeing like a state: How certain schemes to improve the human condition have failed* (New Haven: Yale University Press)

- [7] Susser M and Susser E 1996 *Am J Public Health* **86** 674-77
- [8] Bizouarn P 2016 *Médecine/Sciences* **32** 500-505
- [9] Nersessian N J 2012 *Mind, Culture, and Activity* **19** 222-39
- [10] Berg M and Timmermans S 2000 *Configurations* **8** 31-61
- [11] Latour B 2005 *Reassembling the social* (Oxford, England: Oxford University Press)
- [12] Gorin J and Mislevy R 2013 *Inherent measurement challenges in the next generation science standards for both formative and summative assessment* (Princeton, NJ: ETS)
- [13] Wilson M 2004 *Towards coherence between classroom assessment and accountability* (Chicago: University of Chicago Press)
- [14] National Research Council 2006 *Systems for state science assessment* eds M R Wilson and M W Bertenthal (Washington, DC: The National Academies Press)
- [15] Latour B 2011 *Breakthrough J* **2** 21-8
- [16] Haraway D J 1996 *The disunity of science* eds P Galison and D J Stump (Stanford, California: Stanford University Press) pp 428-41
- [17] Dewey J 2012 *Unmodern philosophy and modern philosophy* (Carbondale, Illinois: Southern Illinois University Press)
- [18] Golinski J 2012 *Osiris* **27** 19-36
- [19] Overton W F 2015 *Handbook of child psychology and developmental science, Volume I* eds W Overton and P C M Molenaar (Hoboken, NJ: Wiley) pp 9-62
- [20] Mari L and Wilson M 2014 *Measurement* **51** 315-27
- [21] Pendrill L and Fisher W P Jr 2015 *Measurement* **71** 46-55
- [22] Pendrill L 2014 *NCSLi Measure: The Journal of Measurement Science* **9** 22-33
- [23] Wilson M, Mari L, Maul A and Torres Irribara D 2015 *J Physics Conf Series* **588**(012034)
- [24] Fisher W P Jr and Stenner A J 2016 *Measurement* **92** 489-96
- [25] Galison P 1997 *Image and logic: A material culture of microphysics* (Chicago: University of Chicago Press)
- [26] Latour B 1993 *We have never been modern* (Cambridge, Massachusetts: Harvard University Press)
- [27] Fisher W P Jr 2010 *J Physics Conf Series* **238**
- [28] Wilson M R 2013 *Measurement* **46** 3766-74
- [29] Rasch G 1980 *Probabilistic models for some intelligence and attainment tests* (Chicago: University of Chicago Press)
- [30] Andrich D 1978 *Psychometrika* **43** 561-73
- [31] Wilson M 2005 *Constructing measures: An item response modeling approach* (Mahwah, New Jersey: Lawrence Erlbaum Associates)
- [32] Stenner A J, Fisher W P Jr, Stone M H and Burdick D S 2013 *Frontiers in Psychology: Quantitative Psychol and Measurement* **4** 1-14
- [33] DeBoeck P and Wilson M 2004 *Explanatory item response models: A generalized linear and nonlinear approach* (New York: Springer-Verlag)
- [34] Mead R J 2009 *J Appl Measurement* **10** 208-24
- [35] Beretvas N and Kamata A 2007 *Rasch measurement: Advanced and specialized applications* ed E V Smith Jr and R M Smith (Maple Grove, MN: JAM Press) pp 290-470
- [36] He W and Kingsbury G G 2016 *J Physics Conf Series* **772** 012022
- [37] Fisher W P Jr and Wilson M 2015 *Pensamiento Educativo: Revista de Investigacion Educativa Latinoamericana* **52** 55-78
- [38] Ricoeur P 1992 *Oneself as another* (Chicago, Illinois: University of Chicago Press)
- [39] Dewey J 1954 *The public and its problems* (Athens, Ohio: Swallow Press, Ohio University Press)
- [40] Hutchins E 2014 *Philosophical Psychol* **27** 34-49
- [41] Durakbasa N M, Bauer J M, Bas G and Riepl D 2015 *Internatl J Online Engineering* **11** 14-9